AMBAR N. SENGUPTA

INTRODUCTORY CALCULUS NOTES

Introductory Calculus Notes

Ambar N. Sengupta

17th November, 2011

Contents

Preface 9
Introduction 10

1 Sets: Language and Notation **13**
1.1 Sets and Elements 13
1.2 Everything from nothing 14
1.3 Subsets 15
1.4 Union, Intersections, Complements 17
1.5 Integers and Rationals 17
1.6 Cartesian Products 18
1.7 Mappings and Functions 19
1.8 Sequences 23

2 The Extended Real Line **25**
2.1 The Real Line 25
2.2 The Extended Real Line 26

3 Suprema, Infima, Completeness **29**
3.1 Upper Bounds and Lower Bounds 29
3.2 Sup and Inf: Completeness 30
3.3 More on Sup and Inf 31

4 Neighborhoods, Open Sets and Closed Sets **33**
4.1 Intervals 33
4.2 Neighborhoods 34
4.3 Types of points for a set 35
4.4 Interior, Exterior, and Boundary of a Set 37
4.5 Open Sets and Topology 38
4.6 Closed Sets 39

4.7 Open Sets and Closed Sets 40
4.8 Closed sets in $\mathbb{R}$ and in $\mathbb{R}^*$ 40

5 Magnitude and Distance 41
5.1 Absolute Value 41
5.2 Inequalities and equalities 42
5.3 Distance 43
5.4 Neighborhoods and distance 43

6 Limits 45
6.1 Limits, Sup and Inf 46
6.2 Limits for $1/x$ 49
6.3 A function with no limits 51
6.4 Limits of sequences 52
6.5 Lim with sups and infs 54

7 Limits: Properties 57
7.1 Up and down with limits 57
7.2 Limits: the standard definition 59
7.3 Limits: working rules 61
7.4 Limits by comparing 64
7.5 Limits of composite functions 66

8 Trigonometric Functions 69
8.1 Measuring angles 69
8.2 Geometric specification of sin, cos and tan 70
8.3 Reciprocals of sin, cos, and tan 74
8.4 Identities 74
8.5 Inequalities 77
8.6 Limits for sin and cos 78
8.7 Limits with $\sin(1/x)$ 79
8.8 Graphs of trigonometric functions 80
8.9 Postcript on trigonometric functions 81
Exercises on Limits 82

9 Continuity 85
9.1 Continuity at a point 85
9.2 Discontinuities 85

9.3 Continuous functions 86
9.4 Two examples using $\mathbb{Q}$ 87
9.5 Composites of continuous functions 88
9.6 Continuity on $\mathbb{R}^*$ 88

10 The Intermediate Value Theorem **91**
10.1 Inequalities from limits 92
10.2 Intermediate Value Theorem 93
10.3 Intermediate Value Theorem: a second formulation 94
10.4 Intermediate Value Theorem: an application 95
10.5 Locating roots 96

11 Inverse Functions **99**
11.1 Inverse trigonometric functions 99
11.2 Monotone functions: terminology 102
11.3 Inverse functions 103

12 Maxima and Minima **107**
12.1 Maxima and Minima 107
12.2 Maxima/minima with infinities 110
12.3 Closed and bounded sets 111

13 Tangents, Slopes and Derivatives **113**
13.1 Secants and tangents 114
13.2 Derivative 116
13.3 Notation 117
13.4 The derivative of x^2 118
13.5 Derivative of x^3 120
13.6 Derivative of x^n 121
13.7 Derivative of $x^{-1} = 1/x$ 121
13.8 Derivative of $x^{-k} = 1/x^k$ 122
13.9 Derivative of $x^{1/2} = \sqrt{x}$ 123
13.10 Derivatives of powers of x 125
13.11 Derivatives with infinities 125

14 Derivatives of Trigonometric Functions **127**
14.1 Derivative of sin is cos 127
14.2 Derivative of cos is $-\sin$ 129

14.3 Derivative of tan is $\sec^2$ 129

15 Differentiability and Continuity 131
15.1 Differentiability implies continuity 131

16 Using the Algebra of Derivatives 133
16.1 Using the sum rule 134
16.2 Using the product rule 134
16.3 Using the quotient rule 135

17 Using the Chain Rule 137
17.1 Initiating examples 137
17.2 The chain rule 138

18 Proving the Algebra of Derivatives 141
18.1 Sums 141
18.2 Products 141
18.3 Quotients 143

19 Proving the Chain Rule 145
19.1 Why it works 145
19.2 Proof the chain rule 146

20 Using Derivatives for Extrema 151
20.1 Quadratics with calculus 152
20.2 Quadratics by algebra 153
20.3 Distance to a line 155
20.4 Other geometric examples 160
Exercises on Maxima and Minima 164

21 Local Extrema and Derivatives 167
21.1 Local Maxima and Minima 167
Review Exercises 169

22 Mean Value Theorem 171
22.1 Rolle's Theorem 171
22.2 Mean Value Theorem 172
22.3 Rolle's theorem on $\mathbb{R}^*$ 174

23 The Sign of the Derivative 175
23.1 Positive derivative and increasing nature 175
23.2 Negative derivative and decreasing nature 179
23.3 Zero slope and constant functions 179

24 Differentiating Inverse Functions 181
24.1 Inverses and Derivatives 182

25 Analyzing local extrema with higher derivatives 185
25.1 Local extrema and slope behavior 185
25.2 The second derivative test 188

26 Exp and Log 191
26.1 Exp summarized 191
26.2 Log summarized 193
26.3 Real Powers 195
26.4 Example Calculations 197
26.5 Proofs for Exp and Log 198

27 Convexity 205
27.1 Convex and concave functions 205
27.2 Convexity and slope 206
27.3 Checking convexity/concavity 208
27.4 Inequalities from convexity/concavity 209
27.5 Convexity and derivatives 213
27.6 Supporting Lines 215
27.7 Convex combinations 218
Exercises on Maxima/Minima , Mean Value Theorem, Convexity 223

28 L'Hospital's Rule 225
28.1 Examples 226
28.2 Proving l'Hospital's rule 228
Exercises on l'Hosptal's rule 232

29 Integration 233
29.1 From areas to integrals 233
29.2 The Riemann integral 235
29.3 Refining partitions 237
29.4 Estimating approximation error 239

29.5 Continuous functions are integrable 240
29.6 A function for which the integral does not exist 242
29.7 Basic properties of the integral 243

30 The Fundamental Theorem of Calculus 245
30.1 Fundamental theorem of calculus 245
30.2 Differentials and integrals 246
30.3 Using the fundamental theorem 249
30.4 Indefinite integrals 252
30.5 Revisiting the exponential function 254

31 Riemann Sum Examples 257
31.1 Riemann sums for $\int_1^N \frac{dx}{x^2}$ 257
31.2 Riemann sums for $1/x$ 260
31.3 Riemann sums for x 261
31.4 Riemann sums for x^2 264
31.5 Power sums 266

32 Integration Techniques 269
32.1 Substitutions 269
32.2 Some trigonometric integrals 276
32.3 Summary of basic trigonometric integrals 280
32.4 Using trigonometric substitutions 282
32.5 Integration by parts 286
Exercises on the Substitution Method 290

33 Paths and Length 291
33.1 Paths 291
33.2 Lengths of paths 294
33.3 Paths and Curves 295
33.4 Lengths for graphs 297

34 Selected Solutions 301
Bibliography 324

Preface

These notes are being written for an introductory honors calculus class, Math 1551, at LSU in the Fall of 2011.

The approach is quite different from that of standard calculus texts. (In fact if I had to choose a subtitle for these notes, it would be 'An Anti-calculus-text Book'.) We use natural, but occasionally unusual, definitions for basic concepts such as limits and tangents. We also avoid several stranger aspects of the universe of calculus texts, such as counterintuitive notions of what counts as 'local maximum' or obsessing over 'convex up/down', and stay with practice that is consistent with the way mathematicians actually work. For most topics we show how to work with the method first and then go deeper into proofs and finer points. We prove several results in sharper formulations than seen in calculus texts. Among drastic departures from the standard approaches we work with the extended real line $\mathbb{R}^* = \mathbb{R} \cup \{-\infty, \infty\}$, and define limits in such a way that no special exceptions need to be made for limits involving $\pm\infty$. We follow a consistent strategy of using suprema and infima, which form a running theme through the historical development of the real line and calculus. An entire chapter is devoted to convexity.

For various corrections and comments I thank Justin Katz.

Introduction

There are two fundamental notions that led to the development of calculus historically: (i) the measurement of areas of curved regions, and (ii) the study of tangents to curves. These apparently disconnected themes, formalized in *integral calculus* and *differential calculus*, respectively, come together in the *fundamental theorem of calculus*, that makes the subject so useful and powerful.

Lengths of line segments are measured by comparing to a chosen 'unit' of length. For areas the natural extension would be to measure the area of a region by counting the number of unit squares (squares with unit-length sides) needed to cover the region exactly. This works very well for a rectangle. Clearly a rectangle whose sides are of length 3 units and 4 units is covered exactly by a 3×4 grid of 12 unit squares. This leads naturally to

$$\text{area of a rectangle} = \text{product of the lengths of its sides.}$$

It requires only a few natural and simple steps to compute the area of a triangle (realize it as made up of two halves of rectangles) and more generally the area of a polygonal region.

This strategy fails when we think of a curved region, such as a disk. There is surely no obvious way to cover a disk exactly with a finite number of squares or pieces of squares. (Whether such slicing and reassembly of regions is really possible, and in what sense, is a truly difficult problem.) The solution to this problem for specific curved regions was already known in the era of Greek mathematics. Consider non-overlapping squares lying inside a disk. Surely the area of the disk is at least as large as the sum of the areas of such squares:

$$\text{areas of polygonal regions inside disk} \leq \text{area of disk.}$$

On the other hand, if we cover the disk with squares, which spill over to the outside of the disk, and add up the areas of these squares we obtain an over estimate of the area of the disk:

$$\text{area of disk } \leq \text{areas of polygonal regions covering disk.}$$

Thus, the area of the disk ought to be

> *the unique value that lies between these overestimates and the underestimates.*

This idea, of pinning down a value by realizing it as being squeezed in between overestimates and underestimates is an enormously powerful idea, running all through the foundations of calculus. We will use this idea persistently in developing the basic notions of both integral calculus and differential calculus.

Returning to the disk, it turns out that there is indeed such a unique value lying between the underestimates and the overestimates. The area of the disk scales up by a factor of 9 if the radius is scaled up by a factor of 3. Indeed, the ratio of the area of the disk to that of the square on the radius is the fundamental constant denoted

$$\pi.$$

The simplest underestimate for π is obtained by inscribing a square in the unit circle so that the diagonal of the square is a diameter of the circle; each side of this square has length $\sqrt{2}$ and so its area is 2. Next, drawing a square with width given by the diameter of the circle gives an overestimate for the area: 2^2. Thus

$$2 < \pi < 4.$$

Archimedes, working with a 96-sided polygon obtained the estimates:

$$3\frac{1}{7} < \pi < 3\frac{10}{71}.$$

Aside from such estimates a crucial point is that the overestimates and underestimates can be made ‘arbitrarily’ close to each other; more precisely, there can only be a *unique* number lying between the overestimates and underestimates.

The value of π expressed as a decimal is:

$$\pi = 3.141\,592\,6\ldots$$

where the dots mean that the decimal continues endlessly. What this means exactly is that π is the least number greater than all the finite decimal approximations listed through the decimal expansion. This expression, though quite concrete, is frustrating in that there is no direct specificaton of, say, what the 25-th decimal entry is. More informative are formulas such as

$$\pi = 4\left[1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots\right],$$

or the enormously efficient but mysterious Ramanujan formula

$$\pi = \frac{9801}{2\sqrt{2}\sum_{n=0}^{\infty} \frac{(4n)!(1103+26390n)}{(n!)^4(396)^{4n}}}.$$

What exactly these formulas mean will become clear when we have studied limits and infinite series sums.

The number π, originating in geometry, appears in a vast array of contexts in physics, chemistry, engineering and statistics.

Archimedes was able to compute areas of more complex curved regions by very careful estimations, encoded in his *method of exhaustion* for computing areas. The great power of integral calculus is illustrated by the fact that it turns the genius of Archimedes' method into an utterly routine calculation that a student of calculus can do in moments.

Chapter 1

Sets: Language and Notation

Set theory provides the standard foundation for nearly all of mathematics. It is, however, an especially abstruse area in mathematics. In this chapter we will learn some language and notation from set theory that is widely used in mathematics.

1.1 Sets and Elements

A *set*, for practical purposes, is a collection of objects. These objects are called the *elements* of the set. For example,

$$\{-1, D, 5, 8\}$$

is a set whose elements are -1, D, 5, and 8. This is a typical way of displaying a set: list its elements, separated by commas, within the braces { and }.

Sometimes a set if best specified by describing its elements. For instance,

$$\{x : x \text{ is a prime number less than } 10\}$$

is the same as the set

$$\{2, 3, 5, 7\}.$$

Occasionally, we will dispense with formality and simply write the descriptive form of the set as

$$\{\text{all primes less than } 10\}.$$

The notation

$$x \in y$$

says that x is an element of y. Thus,

$$3 \in \{2, 3, 5, 8\}.$$

On the other hand 4 is not an element of $\{2, 3, 5, 8\}$, and this is displayed as

$$4 \notin \{2, 3, 5, 8\}.$$

Sets x and y are said to be *equal*, written

$$x = y$$

if every element of x is an element of y and every element of y is an elements of x; in other words, sets are equal if they have the same elements.

Both $\{2, 3, 5, B\}$ and $\{B, 5, B, 2, 3\}$ display exactly the same set. If we happen to repeat elements in one display or if we display the elements in a different order it does not change the set. As another example,

$$\{3, 5\} = \{5, 3\}.$$

1.2 Everything from nothing

The *empty set* $\emptyset$ is the set that contains no elements at all:

$$\emptyset = \{\,\}. \tag{1.1}$$

For the empty set any statement such as

$$a \in \emptyset$$

is false, for $\emptyset$ contains no element.

Notice that the set

$$\{\emptyset\}$$

is *not* empty: it contains one element, that being $\emptyset$. This is a bit confusing, so think it over. Thus:

$$\{\emptyset\} \neq \emptyset \tag{1.2}$$

because the set on the left contains one element whereas the set on the right doesn't contain any element.

In fact, $\{\emptyset\}$ has a name. It is just the mathematical definition of 1:

$$1 \overset{\text{def}}{=} \{\emptyset\}. \tag{1.3}$$

Having 1 and $\emptyset$ we can form another set:

$$\{\emptyset, 1\}.$$

This, of course, is 2:

$$2 \overset{\text{def}}{=} \{\emptyset, 1\}. \tag{1.4}$$

In this way we obtain the numbers

$$0 = \emptyset, 1, 2, 3, \ldots,$$

which, together, form our first *infinite* set:

$$\{0, 1, 2, 3, \ldots\}.$$

Here we have identified 0 and the empty set, but conceptually one thinks of 0 as 'the number of elements' in the empty set rather than $\emptyset$.

Much more can be done. The negative numbers

$$-1,\ -2,\ -3, \ldots$$

can also be constructed as sets, and then the rational numbers, such as $-13/271$. In fact, virtually every object in mathematics is a set.

There is, of course, much more to numbers than simply names given to certain sets, but we will not pursue this direction further. It is also good to keep in mind that our notion of numbers, both for counting and for ordering (such as listing items as first, second, third, etc.) is ancient, whereas the formalizing of this notion within set theory is barely over a hundred years old.

1.3 Subsets

If x and y are sets, we say that x is a *subset* of y if every element of x is an element of y; we denote this by

$$x \subset y.$$

For instance,

$$\{1,2,5,6\} \subset \{0,5,6,4,7,2,3\}.$$

Note that every set is a subset of itself:

$$x \subset x.$$

Here are a couple more simply observations:

- if $x \subset y$ and $y \subset x$ then $x = y$;
- if $x \subset y$ and $y \subset z$ then $x \subset z$.

Sometimes one gets confused between $a \in b$ and $x \subset y$. These are different notions. For example,

$$3 \in \{1,3,5\}$$

but 3 is not a subset of $\{1,3,5\}$. But here is a somewhat twisted example: for the set

$$a = \{1, \{1\}\}$$

we have both

$$\{1\} \in a, \qquad \text{and} \qquad \{1\} \in a.$$

But this is unusual.

One can do strange things with the empty set, always using arguments by contradiction. Here is a starter instance of this:

Proposition 1.3.1 *The empty set is a subset of every set:*

$$\emptyset \subset x \qquad \textit{for all sets } x.$$

Proof. We argue by contradiction. Suppose x is a set and $\emptyset$ is *not* a subset of x. This would mean that $\emptyset$ contains some element that is not in x. But $\emptyset$ contains no element at all. Thus we have a contradiction, and so $\emptyset$ is in fact a subset of x. $\boxed{\text{QED}}$

1.4 Union, Intersections, Complements

The *union* of sets x and y is the set obtained by pooling together their elements into one set. The union is denoted

$$x \cup y.$$

For example,

$$\{1, 3, 5\} \cup \{3, 5, 6, 7\} = \{1, 3, 5, 6, 7\}.$$

One can do unions of any family of sets. For example,

$$\{1\} \cup \{1, 2\} \cup \{1, 2, 3\} \cup \ldots = \{1, 2, 3, \ldots\}.$$

The *intersection* of sets x and y is the set containing the elements that are both in x and in y, and is denoted

$$x \cap y.$$

For example,

$$\{1, 3, 5, 6\} \cap \{2, 4, 6, 3, 8\} = \{3, 6\}.$$

The intersection can be empty of course:

$$\{2, 4, 5\} \cap \{1, 3, 7\} = \emptyset.$$

If the intersection of sets x and y is empty we say that x and y are *disjoint.*

One can take intersections of more than two sets as well:

$$\{1, 5, 3, 6\} \cap \{2, 3, 4\} \cap \{3, 8, 9\} = \{3\}.$$

Sometimes we are working within one fixed big set X. Then the *complement* of any given subset $A \subset X$ is the set of elements of X not in A:

$$A^c = \{p \in X \ : \ p \notin A\}.$$

1.5 Integers and Rationals

The numbers $0, 1, 2, 3, \ldots$ along with their negatives form the set $\mathbb{Z}$ of *integers*:

$$\mathbb{Z} = \{0, 1, -1, 2, -2, 3, -3, \ldots\}. \tag{1.5}$$

Taking ratios of integers yields the *rational numbers*. Thus a rational number can be expressed as

$$p/q,$$

where p and q are integers, with $q \neq 0$ of course. For example, $-34/15$ is rational. The set of all rational numbers is denoted

$$\mathbb{Q} = \{p/q \,:\, p, q \in \mathbb{Z},\, q \neq 0\}. \tag{1.6}$$

As we know there is a whole algebra for the rationals: they can be added, substracted, multiplied, divided (not by zero). Moreover, there is an *order* relation on $\mathbb{Q}$, telling us which of two given rationals is bigger.

For calculus we need the set $\mathbb{R}$ of all *real numbers*. This is a much larger set than $\mathbb{Q}$, as it contains *irrational* numbers such as $\sqrt{2}$ and π, in addition to all the rational numbers.

1.6 Cartesian Products

The displays $\{a, b\}$ and $\{b, a\}$ describe the *same* set, but sometimes we need to express a appearing first followed by b. For this purpose there is the notion of an *ordered pair*:

$$(a, b).$$

There are a number of ways to construct a set out of a and b, reflecting the distinction between them; a simple (though certainly not obvious) strategy is to define

$$(a, b) = \{\{a\}, \{a, b\}\}. \tag{1.7}$$

We can check that, with this definition,

$$(a, b) = (c, d) \text{ if and only if } a = c \text{ and } b = d.$$

Thus, for instance,

$$(1, 3) \neq (3, 1).$$

Next, from sets A and B we can construct the set of all ordered pairs (a, b), drawing the first entry a from A, and the second entry b from B:

$$A \times B = \{(a, b) \,:\, a \in A, \quad b \in B\}. \tag{1.8}$$

This is called the *Cartesian product* of A with B. For example,

$$\{2,5,6\} \times \{d,g\} = \{(2,d),(2,g),(5,d),(5,g),(6,d),(6,g)\}.$$

The Cartesian product of a set A with itself is denoted A^2:

$$A^2 = A \times A. \tag{1.9}$$

Thus the *plane*, coordinatized by real numbers, can be modeled mathematically as

$$\mathbb{R}^2 = \{(x,y) \, : x \in \mathbb{R}, y \in \mathbb{R}\}. \tag{1.10}$$

1.7 Mappings and Functions

In calculus we work with *functions* specified by formulas such as

$$y = x^3 + x^2 + 1.$$

This relation is not read as simply an equality of two quantities y and $x^3 + x^2 + 1$, but rather as a procedure for computing one quantity from the value of another:

$$\text{given the value } x = 2 \text{ we compute } y = 2^3 + 2^2 + 1 = 13.$$

Thus what we have here is a prescription: an input value for x leads to an ouput value y. Of course, the letters x and y are in themselves of no significance; the same function is specified by

$$s = t^4 + t^2 + 1.$$

Sometimes a function is specified not by a formula but by an explicit description; for example,

$$1_{\text{prime}}(m) = \begin{cases} 1 & \text{if } m \text{ is a prime number;} \\ 0 & \text{if } m \text{ is not a prime number.} \end{cases}$$

specifies a 'function of m', where m runs over the positive integers. For example,

$$1_{\text{prime}}(5) = 1, \quad \text{and} \quad 1_{\text{prime}}(4) = 0.$$

To formalize the notion of a function as producing an output value from an input value drawn from some given set in set language, observe that the function is essentially known completely if we are given the value $f(t)$ for every relevant t; this information can be encoded by providing the set

$$\{\big(t, f(t)\big) \,:\, t \text{ running over a given set of interest}\}.$$

Note that to keep things unambiguous, $f(t)$ should mean exactly one *unique* value, and not multiple values. Thus, for example,

$$y = \pm\sqrt{1 - x^2}$$

is a relation containing meaningul information but we will not use the term 'function' for this.

We can now turn to a formal definition that reflects this notion.

A *map*, *mapping*, or *function*

$$f : A \to B$$

is specified by a set A, called the *domain* of f, a set B, called the *codomain* of f, and a set $\mathrm{Gr}(f)$ of ordered pairs (a, b), with $a \in A$ and $b \in B$, such that for each $a \in A$ there is a *unique* $b \in B$ for which $(a, b) \in \mathrm{Gr}(f)$. If $(a, b) \in \mathrm{Gr}(f)$ we denote b by $f(a)$:

$$b = f(a) \text{ means } (a, b) \in \mathrm{Gr}(f).$$

The set $\mathrm{Gr}(f)$ is the graph of the mapping f.

The term *function* is used normally, instead of mapping, in calculus.

In order to make progress we also need to use language shortcuts. For example, instead of saying the function, with domain and codomain the set of real numbers and having graph

$$\{(a, a^2) \,:\, a \in \mathbb{R}\},$$

we often simply write

$$\text{the function } y = x^2,$$

it being clear, unless otherwise spelled out, that x runs over all real numbers.

Thus, when we say

$$\text{the function } y = x$$

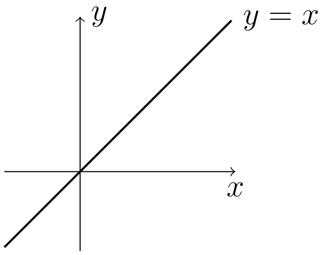


Figure 1.1: Graph of $y = x$

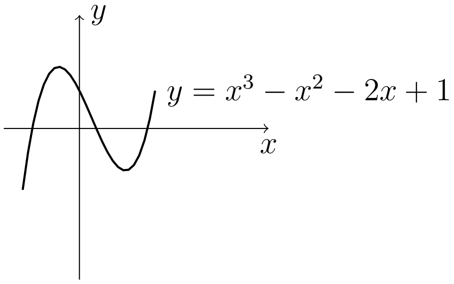


Figure 1.2: Graph of $y = x^3 - x^2 - 2x + 1$

we mean the function x whose value at any p is p itself; for instance,

$$x(3) = 3, \quad x(-4.75) = -4.75.$$

The graph of $y = x$ is displayed visually as a straight line:

The graph of $y = x^3 - x^2 - 2x + 1$ is displayed visually as

Often in calculus, the codomain is clear from context (usually, the set $\mathbb{R}$ of all real numbers) and we identify a function with its graph, ignoring specification of the codomain.

The *range* of a function f is the set of all values it takes:

$$\text{Range}(f) = \{f(a)\ : a \in \text{ domain of } f\ \}.$$

For example, for the function

$$y = x^2 \quad \text{for all real numbers } x,$$

the range is the set of all non-negative real numbers

$$[0, \infty).$$

For the function 1_{prime} we considered before, the range is the set $\{0,1\}$:

$$\text{Range}(1_{\text{prime}}) = \{0,1\}.$$

Returning to our earlier example

$$y = \pm\sqrt{1-x^2},$$

observe that this is equivalent to

$$x^2 + y^2 = 1.$$

This can be viewed as its *graph*:

$$\{(x,y) \in \mathbb{R} : x \in \mathbb{R}, y \in \mathbb{R}\},$$

which is the circle of unit radius, centered at the origin $(0,0)$.

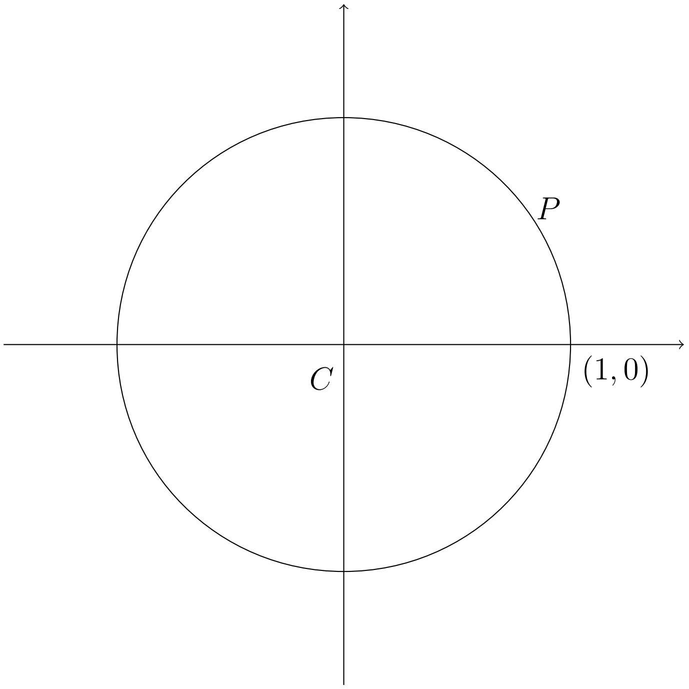


Figure 1.3: Graph of the circle $x^2 + y^2 = 1$.

1.8 Sequences

A function whose domain is the set

$$\mathbb{N} = \{1, 2, 3, \ldots\}$$

of positive integers is called a *sequence*. For example,

$$f : \mathbb{N} \to \mathbb{R} : n \mapsto \frac{1}{n}$$

is a sequence. It is conventional to use n in place of x, and write f_n in place of $f(n)$. In the preceding example we may describe it as

$$\text{the sequence } f_n = 1/n,$$

or, more formally, as

$$\text{the sequence } (f_n)_{n\geq 1} \text{ where } f_n = 1/n,$$

or, most simply, as

$$\text{the sequence } 1/n.$$

Here you have to underdstand from the context that n runs over $\mathbb{N}$, and we are, strictly speaking, talking about the function that associates $1/n$ to every $n \in \mathbb{N}$. The way to *think* about the sequence is to think of it as a list of its values:

$$\frac{1}{1}, \quad \frac{1}{2}, \quad \frac{1}{3}, \quad \frac{1}{4}, \ldots$$

Sometimes there is no 'formula' for the n-th term; for example,

$$\text{the sequence } p_n = n\text{-th prime number.}$$

(There are infinitely many prime numbers and so this does indeed specify a sequence.)

In some contexts it is useful to take the domain of a sequence to include 0 as well. For example, the *factorials*

$$f_n = n! \qquad \text{for all } n \in \{0, 1, 2, \ldots\}$$

are secified by

$$0! = 1, \quad 1! = 1, \quad 2! = 1!2 = 2, \quad 3! = 2!{\cdot}3 = 1{\cdot}2{\cdot}3, \qquad (n{+}1)! = n!{\cdot}(n{+}1).$$

Chapter 2

The Extended Real Line

In this chapter extend the real line $\mathbb{R}$ by including a largest element ∞ and a smallest element $-\infty$. Using these makes it possible to write many theorems in a simpler way, without having a list of qualifiers of which situations need to be excluded.

2.1 The Real Line

The numbers $0, 1, 2, \ldots$ arise from the notion of *counting*. From these it is possible to construct negative numbers $-1, -2, \ldots$ and then the rational numbers.

Geometry leads us beyond the rationals and forces us to bring in other numbers. Consider straightline l, and on it pick a special point O and another point U.

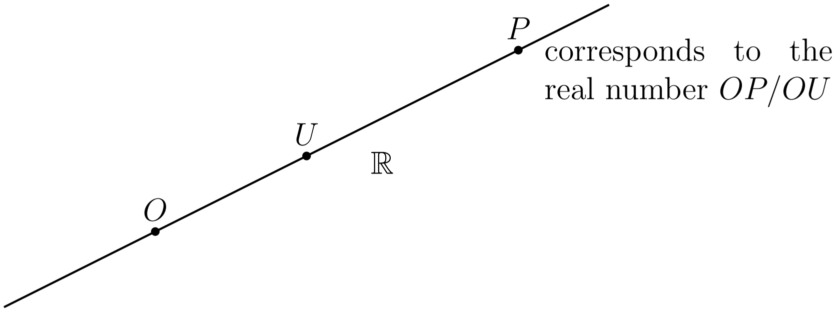


Figure 2.1: Real numbers as ratios of segments

Then for any point P on the line l we can think of the geometical concept of the ratio

$$OP/OU,$$

where we take this to be negative if P is on the opposite side of O from U. Such ratios can be added and multiplied by using geometric constructions (these geometric operations on segments were described in Euclid's *Elements*). Thus they form a system of numbers called the *real numbers*.

For example, if P is just the point U then the ratio

$$OP/OU = OU/OU$$

corresponds to the number 1. Similarly, we have points P for which OP/OU is a rational such as $-4/7$. But there are also points P for which OP/OU cannot be expressed as a ratio of integers.

For example, consider a right angled triangle that has two sides of length OU. Then the diagonal is, by Pythagoras' theorem, has the ratio to OU given by $\sqrt{2}$. It is a fact that $\sqrt{2}$ *is not a rational number*, in that there is no rational number whose square is 2.

The rationals are *dense* in the real line: between any two distinct reals there lies a rational. The irrationals are also dense in the real line: between any two distinct reals lies an irrational.

2.2 The Extended Real Line

The extended real line is obtained by a largest element ∞, and a smallest element $-\infty$, to the real line $\mathbb{R}$:

$$\mathbb{R}^* = \mathbb{R} \cup \{-\infty, \infty\} \tag{2.1}$$

Here ∞ and $-\infty$ are abstract elements. We extend the order relation to $\mathbb{R}$ by declaring that

$$-\infty < x < \infty \qquad \text{for all } x \in \mathbb{R} \tag{2.2}$$

Much of our work will be on $\mathbb{R}^*$, instead of just $\mathbb{R}$.

We define addition on $\mathbb{R}^*$ as follows:

$$\begin{aligned} x + \infty &= \infty = \infty + x \quad \text{for all } x \in \mathbb{R}^* \text{ with } x > -\infty && (2.3)\\ y + (-\infty) &= -\infty = (-\infty) + y \quad \text{for all } y \in \mathbb{R}^* \text{ with } y < \infty. && (2.4) \end{aligned}$$

Note that

$$\infty + (-\infty) \text{ is } \textit{not defined},$$

which just means that there is no useful or consistent definition for it.

For multiplication we set

$$\begin{aligned} x \cdot \infty &= \infty = \infty \cdot x \quad \text{for all } x \in \mathbb{R}^* \text{ with } x > 0 && (2.5)\\ x \cdot (-\infty) &= -\infty = (-\infty) \cdot x \quad \text{for all } x \in \mathbb{R}^* \text{ with } x < 0. && (2.6) \end{aligned}$$

It is a bit dangerous to define $\infty \cdot 0$. However, playing it very carefully, it turns out to be useful to set

$$\begin{gathered} \infty \cdot 0 = 0 \cdot \infty = 0 \\ (-\infty) \cdot 0 = 0 \cdot (-\infty) = 0. \end{gathered} \tag{2.7}$$

This convention is quite useful when we study *integration* theory, but it should not be used in other parts of calculus, such as the study of limits. A definition, aside from being an identification of a distinctive concept, is meaningful and necessary only in so far as it is useful in formulating results. For example, $0/0$ is undefined not because somehow we have arbitrarily decided not to define it, but any definition for it would be a dead end, of no use elsewhere in mathematics.

Some familiar algebraic facts are still valid in $\mathbb{R}^*$:

$$x + y = y + x, \qquad x + (y + z) = (x + y) + z, \tag{2.8}$$

whenever either side of these equations exist (that is we don't have $\infty + (-\infty)$ appearing).

Chapter 3

Suprema, Infima, Completeness

In this chapter we examine a fundamental property of the real line that distinguishes it from the rationals and that makes much of calculus possible. This property is called *completeness*, and roughly it means that the real line has no 'gaps' or 'holes'.

3.1 Upper Bounds and Lower Bounds

Consider a set $S \subset \mathbb{R}^*$.

A point $p \in \mathbb{R}^*$ is an *upper bound* of S if it lies to the right of S:

$$x \leq p \qquad \text{for all } x \in S.$$

For example, for the set

$$\{3, 4, 6\} \cup (8, 9]$$

upper bounds are all numbers ≥ 9. Note that 9 is also an upper bound.

For the entire real line $\mathbb{R}$ the only upper bound is ∞. If we had restricted ourselves to working only with real numbers then $\mathbb{R}$ would not have an upper bound.

Note that ∞ is an upper bound for every subset of $\mathbb{R}^*$.

Here is a mind twister for the empty set: 3 is an upper bound for $\emptyset$. As usual, to prove this we argue by contradiction. Suppose 3 were *not* an upper bound of $\emptyset$. This would mean that 3 is *less* than some $x \in \emptyset$:

$$3 < x \qquad \text{for some } x \in \emptyset.$$

But the empty set has no element and so no such x exists. Thus, 3 must indeed be an upper bound for $\emptyset$.

Is there anything special about 3 in the preceding argument? Definitely not! Thus:

every value in $[-\infty, \infty]$ *is an upper bound of* $\emptyset$.

Now we turn to the flip side of the concept of upper bound. A point $p \in \mathbb{R}^*$ is a *lower bound* of S if it lies to the left of S:

$$p \leq x \qquad \text{for all } x \in S.$$

Thus, for example, for the set

$$(8, 9] \cup [11, \infty)$$

all $p \leq 8$ are lower bounds. Note that 8 is also a lower bound.

The value $-\infty$ is a lower bound for every subset of $\mathbb{R}^*$.

Returning to the strange case of the empty set the same logical gymnastics show that *every value in* $[-\infty, \infty]$ *is a lower bound of* $\emptyset$.

3.2 Sup and Inf: Completeness

Consider any $S \subset \mathbb{R}^*$.

The smallest upper bound of S, that is the *least upper bound* of S, is called the *supremum* of S and denoted

$$\sup S.$$

For example,

$$\sup(8, 9] \cup [11, \infty) = \infty.$$

The largest lower bound, that is the *greatest lower bound*, of S is called its *infimum* and denoted

$$\inf S.$$

With the example above we have

$$\inf(8, 9] \cup [11, \infty) = 8.$$

If you think about the empty set, a strange thing happens. Recall that every value in $\mathbb{R}^*$ is an upper bound of $\emptyset$ and so the least upper bound is $-\infty$:

$$\sup \emptyset = -\infty.$$

Similarly,

$$\inf \emptyset = \infty.$$

Thus, the supremum of the empty set is actually *less* than the infimum!

Here is a fundamental property of $\mathbb{R}^*$:

$$\textit{Every subset of } \mathbb{R}^* \textit{ has a supremum and an infimum.} \tag{3.1}$$

This is called the *completeness* of $\mathbb{R}^*$.

If we want to stay within just the set of real numbers, the statement is a little bit more complicated so as to rule out all the infinities:

If $A \subset \mathbb{R}$ is not empty and has an upper bound then it has a supremum, (3.2)

and an analogous statement holds for the infimum.

The completeness of $\mathbb{R}$ is a crucial property. It does not hold for rationals: roughly speaking if we draw $\mathbb{Q}$ on a line there will be lots and lots of holes, for instance the point where $\sqrt{2}$ would be missing. Many of the great useful results of calculus would fail on $\mathbb{Q}$ just for this reason.

The completeness property can be taken to be an *axiom* in one approach to the study of the real number system. But in another, more constructive ,approach, where $\mathbb{R}$ is constructed out of set theory, completeness is a property that is proved as a fundamental theorem about $\mathbb{R}$.

3.3 More on Sup and Inf

Consider a *non-empty* set $S \subset \mathbb{R}^*$. Pick some point p in S. Then any lower bound of S is $\leq p$ and every upper bound of S is $\geq p$:

$$\text{any lower bound of } S \leq p \leq \text{any upper bound.}$$

In particular,

$$\text{the greatest lower bound of } S \leq p \leq \text{the least upper bound.}$$

Thus,

$$\inf S \leq p \leq \sup S \quad \text{for all } p \in S.$$

If S contains just one point then the inf and sup coincide: for example,

$$\inf\{3\} = 3 = \sup\{3\}.$$

On the other hand

$$\inf S < \sup S \quad \text{if } S \text{ contains more than one point.} \tag{3.3}$$

Now consider another situation. Consider sets

$$B \subset A \subset \mathbb{R}^*.$$

Thus everything in B is also in A. Any upper bound of A is $\geq$ all elements of A and hence is $\geq$ all elements of B. Thus:

$$\text{every upper bound of } A \text{ is an upper bound of } B.$$

In particular,

$$\text{the least upper bound of } A \text{ is an upper bound of } B.$$

In other words:

$$\sup A \text{ is an upper bound of } B.$$

So, of course,

$$\text{the least upper bound of } B \text{ is } \leq \sup A.$$

Thus,

$$\sup B \leq \sup A \qquad \text{if } B \subset A. \tag{3.4}$$

Picking a smaller set decreases the supremum, where smaller means that it is contained in the larger set. ('Decreases' is in a lose sense here, as it may happen that $\sup A$ is equal to $\sup B$.)

By a similar reasoning we have

$$\inf A \leq \inf B \qquad \text{if } B \subset A. \tag{3.5}$$

Picking a smaller set increases the infimum, with qualifiers as before.

Chapter 4

Neighborhoods, Open Sets and Closed Sets

In this chapter we study some useful concepts for studying the concept of nearness of points in $\mathbb{R}^*$.

4.1 Intervals

An *interval* in $\mathbb{R}^*$ is, geometrically, just a segment in the extended real line. For example, all the points $x \in \mathbb{R}^*$ for which $1 \leq x \leq 2$ is an interval. More officially, an interval J is a non-empty subset of $\mathbb{R}^*$ with the property that for any two points of J all points between the two points also lie in J: if $s, t \in J$, with $s < t$, and if $s < p < t$ then $p \in J$.

Let J be an interval, a its infimum and b its supremum:

$$a = \inf J, \quad \text{and} \quad b = \sup J.$$

Consider any point p strictly between a and b. Since $a < p$, the point p is not a lower bound and so there is a point $s \in J$ with $s < p$. Since $b > p$, the point p is not an upper bound, and so there is a point $t \in J$ with $p < t$. Thus

$$s < p < t.$$

Since $s, t \in J$ it follows that p, being between s and t, is also in J. The endpoints a and b themselves might or might not be in J. Thus we have the following possibilities for J:

$$\begin{aligned}
[a,b] &\overset{\text{def}}{=} \{x \in \mathbb{R}^* \,:\, a \leq x \leq b\}\\
[a,b) &\overset{\text{def}}{=} \{x \in \mathbb{R}^* \,:\, a \leq x < b\}\\
(a,b] &\overset{\text{def}}{=} \{x \in \mathbb{R}^* \,:\, a < x \leq b\}\\
[a,b) &\overset{\text{def}}{=} \{x \in \mathbb{R}^* \,:\, a < x < b\}.
\end{aligned} \tag{4.1}$$

An interval of the form $[a, b]$ is called a *closed* interval, and an interval of the form (a, b) is called an *open interval*. Thus a closed interval contains it two endpoints, while an open interval contains neither endpoint.

4.2 Neighborhoods

A *neighborhood* of a point $p \in \mathbb{R}$ is an interval of the form

$$(p-\delta, p+\delta)$$

where $\delta > 0$ is any positive real number. For example,

$$(1.2, 1.8)$$

is a neighborhood of 2.

A typical neighborhood of 0 is of the form

$$(-\epsilon, \epsilon)$$

for any positive real number ϵ.

A *neighborhood of* ∞ in $\mathbb{R}^* = \mathbb{R} \cup \{-\infty, \infty\}$ is a ray of the form

$$(t, \infty] = \{x \in \mathbb{R}^* : x > t\}$$

with t any real number. For example,

$$(5, \infty]$$

is a neigborhood of ∞.

A *neighborhood of* $-\infty$ in $\mathbb{R}^*$ is a ray of the form

$$[-\infty, s) = \{x \in \mathbb{R}^* : x < s\}$$

where $s \in \mathbb{R}$. An example is

$$[-\infty, 4)$$

Observe that *if U and V are neighborhoods of p then $U \cap V$ is also a neighborhood of p.* (In fact, for the type of neighborhoods we have been working with, either U contains V as a subset of vice versa, and so $U \cap V$ is just the smaller of the two neighborhoods.)

Observe also that if N is a neighborhood of a point p, and if $q \in N$ then q had a neighborhood lying entirely inside N. For example, the neighborhood $(2, 4)$ of 3 contains 2.5, and we can form the neighborhood $(2, 3)$ of 2.5 lying entirely inside $(2, 4)$.

Here is a simple but fundamental observation:

$$\textit{Distinct points of } \mathbb{R}^* \textit{ have disjoint neighborhoods.} \tag{4.2}$$

This is called the Hausdorff property of $\mathbb{R}^*$.

For example, 3 and 5 have the neighborhoods

$$(2, 4) \qquad \text{and} \qquad (4.5, 5.5)$$

The points 2 and ∞ have disjoint neighborhoods, such as

$$(-1, 5) \qquad \text{and} \qquad (12, \infty]$$

Exercise Give examples of disjoint neighborhoods of

(i) 2 and -4

(ii) $-\infty$ and 5

(iii) ∞ and $-\infty$

(iv) 1 and -1

4.3 Types of points for a set

Consider a set

$$S \subset \mathbb{R}^*.$$

A point $p \in \mathbb{R}^*$ is said to be an *interior points* of S if it has a neighborhood U lying entirely inside S:

$$U \subset S.$$

For example, for the set

$$E = (-4, 5] \cup \{6, 8\} \cup [9, \infty],$$

the points $-2, 3, 11$ are interior points. The point ∞ is also an interior point of E.

A point p is an *exterior point* if it has a neighborhood U lying entirely outside S:

$$U \subset S^c.$$

For example, for the set E above, points -5, 7, and $-\infty$ are exterior to E.

A point that is neither interior to S nor exterior to S is a *boundary point* of S. Thus p is a boundary point of S if every neighborhood of p intersects both S and S^c.

In the example above, the boundary points of E are

$$-4,\ 5,\ 6,\ 8,\ 9,\ \infty.$$

Next consider the set

$$\{3\} \cup (5, \infty)$$

The boundary points are $3, 5$, and ∞. It is important to observe that if we work with the real line $\mathbb{R}$ instead of the extended line $\mathbb{R}^*$ then we must exclude ∞ as a boundary point, because it doesn't exist as far as $\mathbb{R}$ is concerned.

Example For the set $A = [-\infty, 4) \cup \{5, 9\} \cup [6, 7)$, decide which of the following are true and which false:

(i) -6 is an interior point (T)
(ii) 6 is an interior point (F)
(iii) 9 is a boundary point (T)
(iv) 5 is an interior point (F)

Ecxerise For the set $B = [-\infty, -5) \cup \{2, 5, 8\} \cup [4, 7)$, decide which of the following are true and which false:

(i) -6 is an interior point
(ii) -5 is an interior point
(iii) 5 is a boundary point
(iv) 4 is an interior point
(v) 7 is a boundary point.

4.4 Interior, Exterior, and Boundary of a Set

The set of all interior points of a set S is denoted

$$S^0$$

and is called the *interior* of S.

The set of all boundary points of S is denoted

$$\partial S$$

and is called the *boundary* of S.

The set of all points exterior to S is the *exterior* of S, and we shall denote it

$$S^{\text{ext}}.$$

Thus, the whole extended line $\mathbb{R}^*$ is split up into three disjoint pieces:

$$\boxed{\mathbb{R}^* = S^0 \cup \partial S \cup S^{\text{ext}}} \tag{4.3}$$

Recall that a point p is on the boundary of S if every neighborhood of the point intersects both S and S^c. But this is exactly the condition for p to be on the boundary of S^c. Thus

$$\boxed{\partial S = \partial S^c.} \tag{4.4}$$

The interior of the entire extended line $\mathbb{R}^*$ is all of $\mathbb{R}^*$. So

$$\partial\mathbb{R}^* = \emptyset.$$

Moreover,

$$\partial\emptyset = \emptyset.$$

At another extreme of unexpected behavior is the set $\mathbb{Q}$ of rational numbers. If you take any neighborhood U of any point in $\mathbb{R}^*$ then U contains both rational numbers and irrational numbers. Thus, every point in $\mathbb{R}^*$ is a boundary point of $\mathbb{Q}$:

$$\partial\mathbb{Q} = \mathbb{R}^*.$$

Example For the set $A = [-\infty, 4) \cup \{5, 9\} \cup [6, 7)$,

(i) $A^0 = [-\infty, 4) \cup (6, 7)$

(ii) $\partial A = \{4, 5, 9, 7, 6\}$

(iii) $A^c = [4, 5) \cup (5, 6) \cup [7, 9) \cup (9, \infty]$

(iv) the interior of the complement A^c is

$$(A^c)^0 = (4, 5) \cup (5, 6) \cup (7, 9) \cup (9, \infty]$$

For the set

$$G = (3, \infty)$$

the boundary of G, when viewed as a subset of $\mathbb{R}^*$, is

$$\partial G = \{3, \infty\}.$$

But if we decide to work only inside $\mathbb{R}$ then the boundary of G is just $\{3\}$.

Excerise For the set $B = \{-4, 8\} \cup [1, 7) \cup [9, \infty)$,

(i) $B^0 =$

(ii) $\partial B =$

(iii) $B^c =$

(iv) the interior of the complement B^c is

$$(B^c)^0 =$$

4.5 Open Sets and Topology

We say that a set is *open* if it does not contain any of its boundary points. For example,

$$(2, 3) \cup (5, 9)$$

is open. The set

$$(3, 4]$$

is not open, because it contains 4, which is a boundary point. On the other hand

$$(4, \infty]$$

is open (even though it is not what is usually called ‘an open interval’).

The entire extended line $\mathbb{R}^*$ is open, because it has no boundary points.

The empty set $\emptyset$ is open, because, again, it doesn't have any boundary points.

Notice then that every point of an open set is an interior point. Thus, a set S is open means that

$$S^0 = S.$$

Thus for an open set S each point has a neighborhood contained entirely inside S. In other words, *S is made up of a union of neighborhoods.*

Viewed in this way, it becomes clear that the *union of open sets is an open set.*

It can also be verified that:

The intersection of a finite number of open sets is open.

Exerise Check that the intersection of the sets $(4, \infty)$ and $(-3\,5)$ and $(2, 6)$ is open.

The collection of all open subsets of $\mathbb{R}$ is called the *topology* of $\mathbb{R}$.

The set of all open subsets of $\mathbb{R}^*$ is called the *topology* of $\mathbb{R}^*$.

4.6 Closed Sets

A set S is said to be *closed* if it contains all its boundary points.

In other words, S is closed if

$$\partial S \subset S$$

Thus,

$$[4, 8] \cup [9, \infty]$$

is closed.

But

$$[4, 5)$$

is not closed because the boundary point 5 is not in this set.

The set

$$[3, \infty)$$

is not closed (as a subset of $\mathbb{R}^*$) because the boundary point ∞ is not inside the set. But, viewed as a subset of $\mathbb{R}$ it is closed. So we need to be careful in deciding what is close and what isn't: a set may be closed viewed as a subset of $\mathbb{R}$ but not as a subset of $\mathbb{R}^*$.

The full extended line $\mathbb{R}^*$ is closed.

The empty set $\emptyset$ is also closed.

Note that the sets $\mathbb{R}^*$ and $\emptyset$ are both open and closed.

4.7 Open Sets and Closed Sets

Consider a set $S \subset \mathbb{R}^*$.

If S is open then its boundary points are all outside S:

$$\partial S \subset S^c.$$

But recall that the boundary of S is the same as the boundary of the complement S^c. Thus, for S to be open we must have

$$\partial(S^c) \subset S^c,$$

which means that S^c contains all its boundary points. But this means that S^c is closed.

Thus, *if a set is open then its complement is closed.*

The converse is also true: if a set is closed then its complement is open. Thus,

Theorem 4.7.1 *A subset of $\mathbb{R}^*$ is open if and only if its complement is closed.*

Exercise. Consider the open set $(1,5)$. Check that its complement if closed.

Exercise. Consider the closed set $[4,\infty]$. Show that its complement is open.

4.8 Closed sets in $\mathbb{R}$ and in $\mathbb{R}^*$

The set

$$[3,\infty)$$

is closed in $\mathbb{R}$, but is *not closed* in $\mathbb{R}^*$. This is because in $\mathbb{R}$ it has only the boundary point 3, which it contains; in contrast, in $\mathbb{R}^*$ the point ∞ is also a boundary point and is not in the set. Thus, when working with closed sets it is important to bear in mind the distinction between being closed in $\mathbb{R}$ and being closed in $\mathbb{R}^*$. There is no such distinction for open sets.

Chapter 5

Magnitude and Distance

5.1 Absolute Value

The *absolute value* or *magnitude* of $x^* \in \mathbb{R}$ is the measure of how large x is, without regard to its sign; the absolute value of x is defined by

$$|x| = \begin{cases} x & \text{if } x \geq 0; \\ -x & \text{if } x < 0. \end{cases} \tag{5.1}$$

In the second line $-x$ helps flip the sign of a negative value of x to a positive one:

$$|-3| = -(-3) = 3.$$

Note that

$$|0| = 0$$

and

$$|x| \geq 0 \quad \text{for all } x \in \mathbb{R}^*.$$

Another observation that comes in handy occasionally is:

$$-|x| \leq x \leq |x| \quad \text{for all } x \in \mathbb{R}^*, \tag{5.2}$$

and in fact, of course, x is equal to either $|x|$ (if $x \geq 0$) or $-|x|$ (if $x < 0$). This gives another useful specification of $|x|$:

$|x|$ *is the larger of the numbers* x *and* $-x$.

As a formula this is:

$$|x| = \max\{x, -x\} \quad \text{for all } x \in \mathbb{R}. \tag{5.3}$$

In other words, $|x|$ is x or $-x$, whichever is ≥ 0.

It is clear that

$$|-x| = |x| \qquad \text{for all } x \in \mathbb{R}^*. \tag{5.4}$$

5.2 Inequalities and equalities

If we take two non-negative numbers, say 3 and 5 then we have

$$|3 + 5| = 8 = |3| + |5|.$$

The same works if both numbers are negative:

$$|(-3) + (-5)| = 8 = |-3| + |-5|.$$

But if one is positive and the other negative then the sum of the absolute values wins out over the absolute value of the sum:

$$|5 + (-3)| = 2 \, < \, |5| + |-3|.$$

We can summarize this in the *triangle inequality* for magnitudes:

$$|a + b| \leq |a| + |b|, \tag{5.5}$$

for all $a, b \in \mathbb{R}^*$ excluding, as always, the cases $\infty + (-\infty)$ and $(-\infty) + \infty$. Here is a proof of this: since $|a + b|$ is the larger of $a + b$ and $-(a + b)$, we just need to show that both of these are $\leq$ the sum $|a| + |b|$. For this observe first that

$$a + b \leq |a| + |b| \quad \text{because } a \leq |a| \text{ and } b \leq |b|$$

and then observe that

$$-(a + b) = (-a) + (-b) \leq |a| + |b| \quad \text{because } -a \leq |a| \text{ and } -b \leq |b|.$$

Thus both $a + b$ and $-(a + b)$ are less or equal to $|a| + |b|$, and so the larger of $a + b$ and $-(a + b)$ is $\leq |a| + |b|$. This proves (5.5).

For multiplication we have equality of absolute values:

$$|ab| = |a||b| \tag{5.6}$$

for all $a, b \in \mathbb{R}^*$. You can check this by considering all possible choices of signs for a and b.

5.3 Distance

The *distance* between two real numbers a and b is the magnitude of $a - b$:

$$d(a, b) \overset{\text{def}}{=} |a - b|. \tag{5.7}$$

Here are two basic properties of distance:

$$d(a, a) = 0 \quad \text{for all } a \in \mathbb{R}, \tag{5.8}$$

and

$$\text{if } d(a, b) = 0 \text{ then } a = b. \tag{5.9}$$

There is a third, less obvious, property that is called the *triangle inequality* that is of great use:

$$d(a, c) \leq d(a, b) + d(b, c) \quad \text{for all } a, b, c \in \mathbb{R}. \tag{5.10}$$

This follows from the triangle inequality for mangitudes:

$$d(a, c) = |a - c| + |a - b \,+\, b - c| \leq |a - b| + |b - c| = d(a, b) + d(a, c).$$

The specific measure of distance given by (5.7) is completely natural and intuitive but does not extend nicely to $\mathbb{R}^*$. Other measures of distance can be constructed that work on $\mathbb{R}^*$.

5.4 Neighborhoods and distance

Consider a point $p \in \mathbb{R}$ and a neighborhood of p given by

$$(p - \delta, p + \delta) = \{x \in \mathbb{R} \,:\, p - \delta < x < p + \delta\},$$

where δ is a positive real number. Clearly this neighborhood consists exactly of those points x whose distance from p is less than δ. Thus we have

$$(p - \delta, p + \delta) = \{x \in \mathbb{R} : |x - p| < \delta\} = \{x \in \mathbb{R} \,:\, d(x, p) < \delta\}. \tag{5.11}$$

Chapter 6

Limits

The concept of limit is fundamental to calculus. It is very easy to grasp intuitively but quite difficult to pin down in a completeley precise mathematical way. For example, anyone would agree that x^2 approaches 9 when x approaches 3; but explaining exactly what this means is a subtle matter. In a first run through the theory it may in fact be practical to give up on this precise specification and just rely on intuition. But using out technology of sup and inf makes it somewhat easy to come to grips with the exact meaning of $x^2 \to 4$ asn $x \to 3$.

For the discussions in this chapter and also elsewhere there is some notational care that is needed in working with values $f(x)$ of a function f. Clearly for such quantities, the point x itself must be in the domain of f. Consider a function f with domain $S \subset \mathbb{R}$, and let p be a point in S. If U is a neighborhood of p then part of U might not be inside S, and so when we speak of the values of f on U we need to focus on $f(x)$ for $x \in U \cap S$. For instance, the function f given by

$$f(x) = \sqrt{x} \quad \text{for } x \in [0, \infty)$$

has domain $S = [0, \infty)$, and if we take p to be the point 0 then a typical neighborhood, such as $(-.01, .01)$ of p, falls partly outside S. Putting the restrciction $x \in U \cap S$ makes it possible to talk about the value $f(x)$.

6.1 Limits, Sup and Inf

Consider the function g defined on all real numbers through the formula

$$g(x) = \begin{cases} x^3 & \text{if } x \neq 2 \\ 0 & \text{if } x \neq 2. \end{cases}$$

Intuitively it is clear that as x approaches 2 the value $g(x)$ approaches $2^3 = 8$; *note that the actual value* $g(2)$*, which is given to be* 0*, is irrelevant to this.* We write this symbolically as

$$g(x) \to 8 \quad \text{as } x \to 2$$

or, even more compactly, as

$$\lim_{x \to 2} g(x) = 8.$$

We read this as “$g(x)$ has the limit 8 as x approaches 2’.

Our goal is to pin down the exact meaning of this. For this consider the behavior of the values $g(x)$ when x is restricted to some neighborhood, say $(1.5, 2.5)$ of 2, again ignoring the actual value $g(2)$:

$$\{g(x) \,:\, x \in (1.5, 2.5) \quad \text{and} \quad x \neq 2\}.$$

How high does $g(x)$ get here? Clearly it is

$$\sup\{g(x) \,:\, x \in (1.5, 2.5) \quad \text{and} \quad x \neq 2\} = (2.5)^3 = 15.625$$

and, on the lower side we have

$$\inf\{g(x) \,:\, x \in (1.5, 2.5) \quad \text{and} \quad x \neq 2\} = (1.5)^3 = 3.375.$$

There is a simpler notation for these sups and infs:

$$\sup_{x \in (1.5, 2.5), x \neq 2} g(x) = 15.625$$
$$\inf_{x \in (1.5, 2.5), x \neq 2} g(x) = 3.375.$$

We can improve our understanding of the behavior of $g(x)$ for x approaching 2 by focusing on a smaller neighborhood of 2, say $(1.9, 2.1)$. For this we have

$$\sup_{x \in (1.9, 2.1), x \neq 2} g(x) = 9.261$$
$$\inf_{x \in (1.9, 2.1), x \neq 2} g(x) = 6.859$$

As we have shrunk the neighborhood the supremum has decreases and the infimum has increased. But notice that *the number* 8 (our suspect for the limit) *lies between the sup and the inf* in both cases. Indeed intuition, and in this case easy verification, suggests that:

> the limit $\lim_{x\to 2} g(x)$ always lies between the sup and inf of the values of $g(x)$ on neighborhoods of 2 (always excluding the value $x = 2$).

In fact,

$$\lim_{x\to 2} g(x)$$

is *the unique value* that lies between the sup and inf of the values of $g(x)$ on all neighborhoods of 2 (always excluding the value $x = 2$).

This provides us with an official definition of limit:

Definition 6.1.1 *Let g be a function defined on a set $S \subset \mathbb{R}$ and let p be any point in $\mathbb{R}^*$. We say that $g(x)$ approaches the limit $L \in \mathbb{R}^*$ as $x \to p$, writing this as*

$$\lim_{x\to p} g(x) = L,$$

if L is the unique value that lies between the sups and infs of the values of g in neighborhoods of p (excluding p itself):

$$\inf_{x\in U\cap S, x\neq p} g(x) \le L \le \sup_{x\in U\cap S, x\neq p} g(x) \tag{6.1}$$

The reason for using $x \in U \cap S$ is that $g(x)$ is only be defined for x in the set S. On eother point is that if in fact $U \cap S$ contains no point other than p then the inf and sup in (6.1) are over the empty set and so (6.1) can never hold (for the left side is ∞ and the right side is $-\infty$) for any value of L and so the limit does not exist in this case. Thus there is no possibility of the limit existing if p is not a limit point of S. It is best not to worry about these fine points too much at this stage.

A note of caution: the above definition is not the standard one but is equivalent to it.

Official definitions of the notion of limit are useful in proving theorems but nearly useless in actually computing limits except for very simple functions. We look at two such simple examples now just to make sure the definition produces values in agreement with common sense.

Take for a starter example, the constant function

$$K(x) = 5 \qquad \text{for all } x \in \mathbb{R}.$$

We want to make sure that the official definition 6.1.1 does imply that

$$\lim_{x\to 3} K(x) = 5.$$

To check this consider any neighborhood of 3:

$$(3-\delta, 3+\delta),$$

where δ is any positive real number. Then

$$\sup_{x\in(3-\delta,3+\delta),x\neq 3} K(x) = 5$$

because the set of values $K(x)$ is just $\{5\}$, and also

$$\inf_{x\in(3-\delta,3+\delta),x\neq 3} K(x) = 5.$$

Thus the only value that lies between the sup and the inf is 5 itself, and hence

$$\lim_{x\to 3} K(x) = 5.$$

Now let us move to the function

$$f(x) = x \quad \text{for all } x \in \mathbb{R}.$$

We would like to make sure that Definition 6.1.1 does imply that $\lim_{x\to 6} f(x)$ is 6. Consider the neighborhood

$$(6-\delta, 6+\delta),$$

where δ is a positive real number. Then

$$\{f(x) : x \in (6-\delta, 6+\delta)\} = \{x : x \in (6-\delta, 6+\delta)\},$$

which is just the interval $(6-\delta, 6+\delta)$, but with the point 6 excluded. Hence its sup is $6+\delta$ and its inf is $6-\delta$. What value lies between these two no matter what δ is ? Certainly it is 6:

$$\inf_{x\in(6-\delta,6+\delta),x\neq 6} f(x) \;<\; 6 \;<\; \sup_{x\in(6-\delta,6+\delta),x\neq 6} f(x).$$

Hence,

$$\lim_{x\to 6} f(x) = 6.$$

In fact, it is clear that

$$\lim_{x\to p} x = p,$$

for every $p \in \mathbb{R}^*$. (The case $p = \infty$ or $p = -\infty$ requires a special, but not difficult, argument, because neighborhoods of these points look different from the usual $(p - \delta, p + \delta)$ form.)

Before moving on the fancier explorations here is a warning on notation. There is nothing special about x, which we have been using in writing limits. Instead of $\lim_{x\to p} f(x)$ we could just as well write $\lim_{y\to p} f(y)$ or $\lim_{w\to p} f(w)$:

$$\lim_{x\to p} f(x) = \lim_{y\to p} f(y) = \lim_{w\to p} f(w) = \lim_{\text{blah}\to p} f(\text{blah}).$$

The only rule about notation is that it must be consistent: *never use the same letter to mean two different things in the same equation or statement*!

6.2 Limits for $1/x$

Common sense shows that

$$1/x \to 0, \qquad \text{as } x \to \infty.$$

We will first show that

$$\inf_{x\in(t,\infty)} \frac{1}{x} = 0,$$

for any positive real number t.

Since $1/x > 0$ when x is positive, 0 is a lower bound for such $1/x$ and so the *greatest* lower bound is ≥ 0:

$$\inf_{x\in(t,\infty)} \frac{1}{x} \geq 0.$$

We need only show that this inf cannot be > 0. Denote the inf by b:

$$b = \inf_{x\in(t,\infty)} \frac{1}{x}.$$

Suppose $b > 0$. Now

$$\frac{1}{x} < b$$

whenever $x > 1/b$. So if we take some real number y larger than both $1/b$ and t, say $y = t + 1/b$, then this x is in (t, ∞) and is also $> 1/b$ and so

$$\frac{1}{y} < b.$$

Consequently,

$$\inf_{x \in (t,\infty)} \frac{1}{x} < b.$$

This is impossible, since it is sayng $b < b$. Hence

$$\inf_{x \in (t,\infty)} \frac{1}{x} = 0.$$

This holds for *all* real $t > 0$. Hence the value 0 satisfies

$$\inf_{x \in (t,\infty], x \neq \infty} \frac{1}{x} \leq 0 \leq \sup_{x \in (t,\infty], x \neq \infty} \frac{1}{x}$$

for all positive real t. (It is a tiresome check to see that it holds also for $t \leq 0$, keeing in mind that $x = 0$ is excluded from the domain of $1/x$.) Hence

$$\lim_{x \to \infty} \frac{1}{x} = 0. \tag{6.2}$$

A similar argument shows that

$$\lim_{x \to -\infty} \frac{1}{x} = 0. \tag{6.3}$$

There are two more limits associated with $1/x$. Taking $1/x$ only on the domain of positive values of x we have

$$\lim_{x \to 0, x > 0} \frac{1}{x} = \infty, \tag{6.4}$$

and focusing on negative values we have

$$\lim_{x \to 0, x < 0} \frac{1}{x} = -\infty. \tag{6.5}$$

These are usually written as

$$\begin{aligned}\lim_{x\to 0+}\frac{1}{x}&=\infty\\ \lim_{x\to 0-}\frac{1}{x}&=-\infty.\end{aligned}\tag{6.6}$$

The limit of $1/x$ as $x \to 0$ does not exist. You can check that in any neghborhood of 0, excluding the value $x = 0$ itself, the sup of $1/x$ is ∞ whereas the inf of $1/x$ is $-\infty$, and so there is no *unique* value between these two extremes.

6.3 A function with no limits

Recall the set $\mathbb{Q}$ of all *rational numbers*:

$$\mathbb{Q} = \{\text{all rationals}\}.$$

This is *dense* in $\mathbb{R}$:

every open interval in $\mathbb{R}$ contains rational points.

The same is true of the *irrationals*:

every open interval in $\mathbb{R}$ contains irrational points.

Consider the *indicator function* of $\mathbb{Q}$, taking the value 1 on rationals and 0 on irrationals:

$$1_{\mathbb{Q}}(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q}; \\ 0 & \text{if } x \notin \mathbb{Q}; \end{cases}\tag{6.7}$$

If you take any $p \in \mathbb{R}$ and any neigborhood U of p it is clear that

$$\sup_{x\in U, x\neq p} 1_Q(x) = 1 \quad \text{and} \quad \inf_{x\in U, x\neq p} 1_Q(x) = 0.$$

Thus there can be no *unique* value between these sups and infs, and so

$\lim_{x\to p} 1_Q(x)$ *does not exist for any* $p \in \mathbb{R}$.

6.4 Limits of sequences

Recall that a sequence s is a function whose domain is the set $\mathbb{N} = \{1, 2, 3, \ldots\}$ (and occasionally, we permit 0 in the domain). The value of the function s on the number $n \in \mathbb{N}$ is denoted by

$$s_n$$

rather than $s(n)$, and the function itself is generally written as

$$(s_n)_{n\geq 1}$$

rather than s.

Our definition of the limit $\lim_{x\to p} f(x)$ makes sense for any function f with domain some subset S of $\mathbb{R}$ and with p being a limit point of S. Thus we can apply to the case of sequences, with S being $\mathbb{N}$ and $p = \infty$. For a sequence $(s_n)_{n\geq 1}$ we are interested in the limit

$$\lim_{n\to\infty} s_n.$$

For example,

$$\lim_{n\to\infty} \frac{1}{n} = 0.$$

It is easy to believe that

$$\lim_{n\to\infty} 2^n = \infty. \tag{6.8}$$

One way to prove this officially is by using the inequality

$$2^n > n \qquad \text{for all } n \in \{1, 2, 3, \ldots\}. \tag{6.9}$$

(2^n is the total number of subsets that an n-element set, such as $\{1, 2 \ldots, n\}$, has, and clearly this is more than n itself since each of the n elements itself provides a subset.)

As a consequence of (6.8) we have

$$\lim_{N\to\infty} \frac{1}{2^N} = 0.$$

Sometimes we work with *infinite series sums*: for a sequence $(s_n)_{n\geq 1}$ the *series sum*

$$\sum_{n=1}^{\infty} s_n$$

is defined to be

$$\sum_{n=1}^{\infty} s_n = s_1 + s_2 + \cdots = \lim_{N\to\infty} S_N, \tag{6.10}$$

where S_N is the N-th *partial sum*

$$S_N = s_1 + \cdots + s_N.$$

When working with series, very often the index 0 is allowed and one works with sums

$$\sum_{n=0}^{\infty} s_n.$$

We will not explore these ideas much further at this point. But before moving on let us work out one example. We will work out the value of the infinite series sum

$$1 + \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \cdots$$

(If you think about it, this surely should add up to 2 ... we will verify this intuition using mathematics.)

In the summation notation this is displayed as

$$\sum_{n=0}^{\infty} \frac{1}{2^n}.$$

Let us first work out the partial sum

$$S_N = 1 + \frac{1}{2} + \cdots + \frac{1}{2^N}$$

The clever trick at this stage is to multiply this by 1/2:

$$\frac{1}{2}S_N = +\frac{1}{2} + \cdots + \frac{1}{2^N} + \frac{1}{2^{N+1}}.$$

Observe that this is very similar to S_N itself: when we subtract nearly everything cancels out:

$$S_N - \frac{1}{2}S_N = 1 - \frac{1}{2^{N+1}}.$$

Thus,

$$\left(1 - \frac{1}{2}\right) S_N = 1 - \frac{1}{2^{N+1}}.$$

Now we have the expression for S_N:

$$1+\frac{1}{2}+\cdots+\frac{1}{2^N}=\frac{1-\frac{1}{2^{N+1}}}{1-\frac{1}{2}}. \tag{6.11}$$

Letting $N\to\infty$ produces the infinite series sum

$$1+\frac{1}{2}+\cdots=\frac{1}{1-\frac{1}{2}}=2. \tag{6.12}$$

6.5 Lim with sups and infs

In this section we are going to push our grasp on sup's and inf's to the limit!

Consider a function f on a set $S\subset\mathbb{R}$ and a point $p\in\mathbb{R}^*$. To avoid unimportant technicalities let us assume that p is a limit point of S. We know of then that

$$\inf_{x\in U\cap S, x\neq p} f(x)\leq \sup_{x\in U\cap S, x\neq p} f(x).$$

Now consider another neighborhood V of p. Both U and V contain the neighborhood

$$W=U\cap V.$$

Hence

$$\begin{aligned}\inf_{x\in U\cap S, x\neq p} f(x)&\leq \inf_{x\in W\cap S, x\neq p} f(x)\\ \sup_{x\in W\cap S, x\neq p} f(x)&\leq \sup_{x\in V\cap S, x\neq p} f(x),\end{aligned} \tag{6.13}$$

because shrinking a set raises the inf and lowers the sup. As a consequence we have

$$\inf_{x\in U\cap S, x\neq p} f(x)\leq \sup_{x\in V\cap S, x\neq p} f(x). \tag{6.14}$$

Thus the sup over *any* neighborhood of p is $\geq$ the inf over *all* neighborhoods of p.

Thus, *there always exists a point that lies between the sups and infs of the values* $f(x)$ for x in neighborhoods of p, excluding $x=p$. If there is a *unique* such point then that point is $\lim_{x\to p} f(x)$.

We can reformulate the relation (6.14) as follows: the sup of f over any neigborhood of p is an upper bound for the set of all the inf values of f over all neighborhoods of p. Hence: *the sup of f over any neighborhood V of p is $\geq$ the sup of the values* $\inf_{x\in U\cap S, x\neq p} f(x)$ *for all neighborhoods U of p*:

$$\sup_U \inf_{x\in U\cap S, x\neq p} f(x) \leq \sup_{x\in V\cap S, x\neq p} f(x) \tag{6.15}$$

where $\sup_U$ is the sup over all neighborhoods U of p. Since this holds for all neighborhoods V of p we then conclude that

$$\sup_U \inf_{x\in U\cap S, x\neq p} f(x) \leq \inf_V \sup_{x\in V\cap S, x\neq p} f(x) \tag{6.16}$$

where $\inf_V$ is the inf over all neighborhoods V of p.

Pondering (6.16) we observe that $L = \lim_{x\to p} f(x)$ exists if and only if the two extremes $\sup_U \inf_{x\in U\cap S, x\neq p} f(x)$ and $\inf_V \sup_{x\in V\cap S, x\neq p} f(x)$ are equal, and this common value must be L itself.

To summarize:

Proposition 6.5.1 *Let f be a function defined on a set $S \subset \mathbb{R}$, and p a limit point of S. Then* $\lim_{x\to p} f(x)$ *exists if and only if* $\sup_U \inf_{x\in U\cap S, x\neq p} f(x)$ *and* $\inf_V \sup_{x\in V\cap S, x\neq p} f(x)$ *are equal, and then*

$$\sup_U \inf_{x\in U\cap S, x\neq p} f(x) = \lim_{x\to p} f(x) = \inf_V \sup_{x\in V\cap S, x\neq p} f(x). \tag{6.17}$$

Chapter 7

Limits: Properties

In this chapter we explore the concept of limit and get familiar with some basic properties that make the computation of limits a routine process for many ordinary functions.

7.1 Up and down with limits

Suppose f is a function on $\mathbb{R}$ and

$$\lim_{x\to 3} f(x) = 5.$$

Our intuition suggests that when x is near enough 3 (excluding $x = 3$ itself) the value $f(x)$ should be > 4. Let us see how we can deduce this from the official definition of limit.

Since $f(x) \to 5$ as $x \to 3$, the number 5 is the *unique* value that lies between the sups and infs of the values of $f(x)$ for x (excluding $x = 3$) in any neighborhood of 3. Thus, for instance, 4 does *not* lie between these sups and infs. This means that there is some neighborhood, say U, of 3 such that 4 is not between the sup and the inf of f over U, excluding $f(3)$:

$$4 \notin \left[\inf_{x\in U, x\neq 3} f(x)\,,\ \sup_{x\in U, x\neq 3} f(x)\right].$$

Note that this interval *does* contain 5 because that is actually the limit. So 4 must lie *below* this interval:

$$4 < \inf_{x\in U, x\neq 3} f(x).$$

This means that *all the values of* f *on the neighborhood* U *of* 3, *excluding* $f(3)$ *itself, are* > 4. This is exactly what we had conjectured based on common sense intuition about limits.

If you look over the preceding discussion you see that what makes the argument work is simply that 4 is a value that is $<$ than the limit 5. Thus what we have really proved is this:

Proposition 7.1.1 *Suppose* f *is a function on some subset* $S \subset \mathbb{R}$, *and*

$$L = \lim_{x \to p} f(x),$$

where $p \in \mathbb{R}^*$. *If* b *is any value below* L, *that is* $b < L$ *then there is a neighborhood* U *of* p *on which*

$$f(x) > b \text{ for all } x \in U \cap S \text{ except possibly for } x = p.$$

We have had to write $x \in U \cap S$, and not just $x \in U$, because $f(x)$ might not be defined for all x in U.

It is important not to get bogged down in the notation used: keep in mind the essense of the idea. What we are saying is, in ordinary rough and ready language, if $f(x) \to L$ as $x \to p$ then the values $f(x)$ lie *above* b when x is near p (but not p itself), for any given value $b < L$.

Of course, we can do the same for values above the limit:

Proposition 7.1.2 *Suppose* f *is a function on some subset* $S \subset \mathbb{R}$, *and*

$$L = \lim_{x \to p} f(x),$$

where $p \in \mathbb{R}^*$. *If* u *is any value* $> L$ *then there is a neighborhood* U *of* p *on which*

$$f(x) < u \text{ for all } x \in U \cap S \text{ except possibly for } x = p.$$

We can even put thee two observations together:

Proposition 7.1.3 *Suppose* f *is a function on some subset* $S \subset \mathbb{R}$, *and*

$$L = \lim_{x \to p} f(x),$$

where $p \in \mathbb{R}^*$. *If* u *is any value* $> L$ *and* b *is any value* $< L$ *then there is a neighborhood* U *of* p *on which*

$$b < f(x) < u \text{ for all } x \in U \cap S \text{ except possibly for } x = p.$$

7.2 Limits: the standard definition

Proposition 7.1.3 can be recast and slightly broadened into the following characterization of the notion of limit:

Proposition 7.2.1 *Let f be a function on $S \subset \mathbb{R}$ and $p \in \mathbb{R}^*$. Suppose $L = \lim_{x\to p} f(x)$ exists . Then for any neighborhood W of L there is a neighborhood U of p such that*

$$f(x) \in W \text{ for all } x \in U \cap S, \text{ excluding } x = p.$$

Proof. Suppose first that L is a real number, and not ∞ or $-\infty$. Then in the neighborhood W there is an open interval

$$(b, u),$$

centered at L, where $b, u \in \mathbb{R}$ and $b < u$. Then by Proposition 7.1.3 there is a neighborhood U of p for which

$$b < f(x) < u \text{ for all } x \in U \cap S \text{ except possibly for } x = p.$$

Thus, for all $x \in U \cap S$, except $x = p$, the value $f(x)$ lies in the interval (b, u) and hence in W.

Now consider the case $L = \infty$. A neighborhood of L then contains an interval of the form

$$(b, \infty],$$

for some real number b. By Proposition 7.1.1 there is a neighborhood U of p for which

$$f(x) > b \text{ for all } x \in U \cap S \text{ except possibly for } x = p.$$

Thus, for all $x \in U \cap S$, except $x = p$, the value $f(x)$ lies in the interval $(b, \infty]$ and hence in W.

Lastly, consider the case $L = -\infty$. A neighborhood of L contains an interval of the form

$$[-\infty, u).$$

By Proposition 7.1.2 there is a neighborhood U of p for which

$$f(x) < u \text{ for all } x \in U \cap S \text{ except possibly for } x = p.$$

Thus, for all $x \in U \cap S$, except $x = p$, the value $f(x)$ lies in the interval $[-\infty, u)$ and hence in W. $\boxed{\text{QED}}$

The result can also be run in reverse:

Proposition 7.2.2 *Let f be a function on $S \subset \mathbb{R}$ and let $p \in \mathbb{R}^*$ be a limit point of S. Suppose $L \in \mathbb{R}^*$ has the property that for any neighborhood W of L there is a neighborhood U of p such that*

$$f(x) \in W \textit{ for all } x \in U \cap S, \textit{ excluding } x = p. \tag{7.1}$$

Then $L = \lim_{x \to p} f(x)$.

Proof. Suppose first that L is a real number, and not ∞ or $-\infty$. We will show that L is the unique value that lies between the sups and infs of $f(x)$ for x in all neighborhoods of p (excluding the the value at $x = p$). Suppose this were not true. Suppose U is a neighborhood of p such that

$$\inf_{x \in U \cap S, x \neq p} f(x) > L. \tag{7.2}$$

Pick any real number u between these two values:

$$L < u < \inf_{x \in U \cap S, x \neq p} f(x). \tag{7.3}$$

Consider then the neighborhood of L given by

$$W = (b, u),$$

where $b < L$ and L is the center of the interval W. Then by the condition (7.1) there is a neighborhood U of p for which

$$b < f(x) < u \text{ for all } x \in U \cap S \text{ except possibly for } x = p.$$

But this is impossible since, by (7.3), $u \leq f(x)$ for all $x \in U \cap S$ with $x \neq p$ (the assumtion that p is not an isolated point or an exterior point for S guarantees that $U \cap S$ actually contains a point other than p). This contradiction shows that (7.2) is false, and so

$$\inf_{x \in U \cap S, x \neq p} f(x) \leq L.$$

By a similar argument it also follows that

$$L \leq \sup_{x \in U \cap S, x \neq p} f(x).$$

Thus L lies between all the sups and infs for f as required.

But we still need to show that L is the *unique* such value. Consider any value $L' \neq L$. Suppose first $L' > L$. Consider a neighborhood W of L that does not contain L'. This means that L' is $>$ than all the values in W. By (7.1) there is a neighborhood U of p such that $f(x) \in W$ for all $x \in U \cap S$ with $x \neq p$. Hence for all such values x we have $f(x) < L'$, and so L' does *nt* lie below $\sup_{x \in U\cap S, x\neq p} f(x)$. Thus L' cannot be the limit of $f(x)$ as $x \to p$. By just a similar argument no value $< L$ could be the limit of $f(x)$ as $x \to p$.

The cases $L = \infty$ and $L = -\infty$ are settled by different but similar arguments, just keeping in mind that the neighborhoods of ∞ and $-\infty$ are 'one sided' rays. $\boxed{\text{QED}}$

Because of the preceding two results the definition of limit has the following equivalent formulation (the standard one):

Definition 7.2.1 *Let f be a function on a set $S \subset \mathbb{R}$, and p be any limit point of S. A value $L \in \mathbb{R}^*$ is said to be the limit of $f(x)$ as $x \to p$ if for any neighborhood W of L there is a neighborhood U of p such that $f(x) \in W$ for all $x \in U \cap S$ with $x \neq p$.*

7.3 Limits: working rules

It is an exhausting and largely pointless task to try to use the definition of limit directly in computing actual limits such as

$$\lim_{x\to 3} \frac{x^2 - 9}{x - 3}.$$

It is far more efficient to develop some working rules, basic results, using which more complicated limits can be reduced to simpler ones and then these worked out directly.

There really are just two limits we have worked out directly from the definition:

$$\begin{aligned} \lim_{x\to p} K &= K \\ \lim_{x\to p} x &= p \end{aligned} \tag{7.4}$$

for all constants $K \in \mathbb{R}$ and points $p \in \mathbb{R}^*$.

The first computationally useful result for limits is simply that the limit of a sum is the sum of the limits:

$$\lim_{x\to p} [f(x) + g(x)] = \lim_{x\to p} f(x) \ + \ \lim_{x\to p} g(x). \tag{7.5}$$

There are some qualifiers: we need to assume that the two limits on the right exist and that the sum on the right is defined (it isn't of the form $(-\infty)+\infty$ or $\infty+(-\infty)$). Here is a formal statement:

Proposition 7.3.1 *Suppose f and g are functions on $S \subset \mathbb{R}$, and suppose the limits $\lim_{x\to p} f(x)$ and $\lim_{x\to p} g(x)$ exist, where p is some point in $\mathbb{R}^*$. Assume furthermore that*

$$\{\lim_{x\to p} f(x)\,,\ \lim_{x\to p} g(x)\} \neq \{\infty,\, -\infty\}. \tag{7.6}$$

Then $\lim_{x\to p}[f(x)+g(x)]$ exists and

$$\lim_{x\to p}\,[f(x)+g(x)] = \lim_{x\to p} f(x)\ +\ \lim_{x\to p} g(x). \tag{7.7}$$

The point of the condition (7.6) is to ensure that the sum on the right in (7.7) exists.

We will not prove this result here. Instead we march on to the next result, focused on multiplication:

Proposition 7.3.2 *Suppose f and g are functions on $S \subset \mathbb{R}$, and p is some point in $\mathbb{R}^*$. Then the limit $\lim_{x\to p}[f(x)g(x)]$ exists and*

$$\lim_{x\to p}\,[f(x)g(x)] = \left(\lim_{x\to p} f(x)\right)\left(\lim_{x\to p} g(x)\right), \tag{7.8}$$

provided the two limits on the right and their product exist and this product is not of the form $0\cdot(\pm\infty)$ or $(\pm\infty)\cdot 0$. More clearly, the condition is that the limits $\lim_{x\to p} f(x)$ and $\lim_{x\to p} g(x)$ exist and

$$\begin{aligned} &\{\lim_{x\to p} f(x)\,,\ \lim_{x\to p} g(x)\} \neq \{0,\infty\} \\ \textit{and}\quad &\{\lim_{x\to p} f(x)\,,\ \lim_{x\to p} g(x)\} \neq \{0,\, -\infty\}. \end{aligned} \tag{7.9}$$

It would have been easier to state this result had we never defined $0\cdot\infty$ and $\infty\cdot 0$ as 0, and instead left such products as undefined. But the convention $0\cdot\infty = 0$ is very convenient for integration theory and so we hold on to it.

Here is a quick application of the preceding rules about limits: we can compute the limit of $x^2 + 3x + 4$ as $x \to 1$:

$$\begin{aligned}\lim_{x\to 1}\left(x^2+3x+4\right) &= \lim_{x\to 1} x^2 + \lim_{x\to 1}(3x+4) \quad \text{(if this exists)}\\ &= \left(\lim_{x\to 1} x\right)\left(\lim_{x\to 1} x\right) + \left[\left(\lim_{x\to 1} 3\right)\left(\lim_{x\to 1} x\right) + \lim_{x\to 1} 4\right] \quad \text{(if this exists)}\\ &= 1\cdot 1 + [3\cdot 1 + 4] \qquad \text{(and this does exist!)}\\ &= 8.\end{aligned} \tag{7.10}$$

This is the kind of reasoning one should go through once but it is clearly so routine that it is not worth mentioning all the steps every time. For

$$\lim_{x\to 2}(x^3 + 5x - 2) = 8 + 10 - 2,$$

not only is the result (the value of the limit) perfectly obvious from common sense it is also perfectly obvious how to use the rules of limits to actually prove that the limit is indeed the value stated above.

Going beyond multiplication we consider ratios. Note that a ration

$$\frac{a}{b}$$

is *not* meaningful if the denominator b is 0 or if a and b are both $\pm\infty$. It is useful, for the purposes of the next result to define

$$\frac{a}{\infty} = 0 = \frac{a}{-\infty} \qquad \text{if } a \in \mathbb{R}. \tag{7.11}$$

Proposition 7.3.3 *Suppose f and g are functions on $S \subset \mathbb{R}$, and p is some point in $\mathbb{R}^*$. Then the limit $\lim_{x\to p}\frac{f(x)}{g(x)}$ exists and*

$$\lim_{x\to p}\frac{f(x)}{g(x)} = \frac{\lim_{x\to p} f(x)}{\lim_{x\to p} g(x)}, \tag{7.12}$$

provided the two limits on the right and their ratio exist (this means that the ratio must not look like something/0 *or* $\pm\infty/\pm\infty$*).*

As a simple illustration of what can go wrong in using this let us look at

$$\lim_{x\to 3}\frac{x^2-9}{x-3}.$$

If we just do the ratio of the limits we end up with 0/0, and this is just a case where *the preceding result cannot be applied*. Thus we need to be less lazy and observe that

$$\frac{x^2-9}{x-3} = \frac{(x-3)(x+3)}{x-3} = x+3,$$

from which it is clear that

$$\lim_{x\to 3}\frac{x^2-9}{x-3} = 6.$$

7.4 Limits by comparing

Sometimes we can find the limit of a function by comparing it with other functions that are easier to understand.

The so called 'squeeze theorem' is a case of this. Suppose f, g, and h are functions on a set $S \subset \mathbb{R}$ and $p \in \mathbb{R}^*$ is such that

$$f(x) \le h(x) \le g(x) \tag{7.13}$$

for all x in S that lie in some neighborhood U of p, excluding $x = p$. Assume that $\lim_{x\to p} f(x)$ and $\lim_{x\to p} g(x)$ exist and are equal:

$$L = \lim_{x\to p} f(x) = \lim_{x\to p} g(x).$$

Then $h(x)$, squeezed in between $f(x)$ and $g(x)$, is forced to also approach the same limit L.

Here is a formal statement and proof:

Proposition 7.4.1 *Suppose f, g, and h are functions on a set $S \subset \mathbb{R}$, and $p \in \mathbb{R}^*$ is such that*

$$f(x) \le h(x) \le g(x) \tag{7.14}$$

for all x in S that lie in some neighborhood of p, excluding $x = p$. Assume also that $\lim_{x\to p} f(x)$ and $\lim_{x\to p} g(x)$ exist and are equal:

$$L = \lim_{x\to p} f(x) = \lim_{x\to p} g(x).$$

Then $\lim_{x\to p} h(x)$ exists and is equal to L.

Proof. We will use the second formulation of the definition of limit, given in Definition 7.2.1. (Since the limit $\lim_{x\to p} f(x)$ is assumed to exist, the point p is a limit point of S.)

Consider any neighborhood W of L. Note that W is an interval containing L.

Since $\lim_{x\to p} g(x) = L$, Definition 7.2.1 implies that there is a neigborhood U_1 of p such that $g(x) \in W$ for all $x \in U_1 \cap S$ with $x \neq p$.

Similarly, there is also a neighborhood U_2 of p such that $g(x) \in W$ for all $x \in U_2 \cap S$ with $x \neq p$. Consider then

$$U = U_1 \cap U_2,$$

which is also a neighborhood of p, contained inside both U_1 an U_2. If we take any $x \in U \cap S$, with $x \neq p$, then both $g(x)$ and $f(x)$ lie inside W, and so anything between $g(x))$ and $f(x)$ also lies in W. Hence

$$f(x) \in W$$

for all $x \in U \cap S$, with $x \neq p$. This proves that

$$\lim_{x\to p} f(x) = L. \quad \boxed{\text{QED}}$$

Here is a consequence that is easier to apply sometimes:

Proposition 7.4.2 *Suppose f is a function on a set $S \subset \mathbb{R}$ and $p \in \mathbb{R}^*$ is such that*

$$\lim_{x\to p} |f(x)| = 0.$$

Then

$$\lim_{x\to p} f(x) = 0.$$

Note that we cannot draw any conclusion if we known that $|f(x)| \to 5$, some other nonzero value, because in that case $f(x)$ could fluctuate up and down between 5 and -5.

Proof. Any $a \in \mathbb{R}$ is either equal $|a|$ or to $-|a|$ (if $a < 0$), and we can certainly write

$$-|a| \leq a \leq |a|.$$

Hence

$$-|f(x)| \leq f(x) \leq |f(x)|$$

for all x in S. As $x \to p$ both $|f(x)| \to 0$ and $-|f(x)| \to 0$, and so, by the 'squeeze' theorem, $f(x) \to 0$ as well. $\boxed{\text{QED}}$

Here is a quick application:

$$\lim_{x\to 0} x^3 \sin\left(x + \frac{1}{x}\right) = 0.$$

This follows by using the fact that $|\sin(\cdot)|$ is ≤ 1, which shows that

$$0 \leq \left| x^3 \sin\left(x + \frac{1}{x}\right)\right| \leq |x|^3,$$

to which we can apply Proposition 7.4.2.

7.5 Limits of composite functions

Suppose f and g are functions. The *composite* $f \circ g$ is specified by

$$(f \circ g)(x) = f\big(g(x)\big),$$

and its domain is the set of all x for which this exists.

For example,

$$x \mapsto \sqrt{1 - x^2}$$

is the composite of the function $x \mapsto 1 - x^2$ (for all $x \in \mathbb{R}$) and the function $u \mapsto \sqrt{u}$ (for $u \geq 0$); its domain is $[-1, 1]$.

Turning to limits, it seems clear that

$$\text{if } g(x) \to q, \text{ as } x \to p, \text{ and } f(v) \to L, \text{ as } v \to q,$$

we should be able to conclude that

$$f(g(x)) \to L \text{ as } x \to p.$$

This is an extremely useful method and mostly we use it without even noticing; for example, using the simple result

$$\lim_{w\to 1} \frac{w^3 - 1}{w - 1} = \lim_{w\to 1} \frac{(w-1)(w^2 + w + 1)}{w - 1} + \lim_{w\to 1}(w^2 + w + 1) = 3,$$

we obtain a limit that is at first less obvious:

$$\lim_{x\to 1}\frac{x^{1/3}-1}{x-1}=\lim_{w\to 1}\frac{w-1}{w^3-1}=\frac{1}{3},$$

by using the 'substitution' $x = w^3$.

This type of reasoning can encounter a rare breakdown. As an extreme example, consider the functions F and G given by

$$F(v)=\begin{cases}1 & \text{for } v\neq 0;\\ 2 & \text{if } v=0,\end{cases}$$

and $G(x) = 0$ for all x. Then

$$G(x)\to 0 \text{ as } x\to 0, \text{ and } F(v)\to 1 \text{ as } v\to 0,$$

but $F(G(x))$ is stuck at the value 2 and so

$$F(G(x))\not\to 1 \text{ as } x\to 0.$$

What has gone wrong is that we have rigged the inner function G to keep hitting (in fact it is stuck at) the forbidden' point $v = 0$ which is excluded from consideration when defining $\lim_{v\to 0} F(v)$.

Treading around this obstacle we can formulate the composite limit very delicately in the following result. Recall that a point p is exterior to a set B if p has a neighborhood leying entirely outside B.

Proposition 7.5.1 *Let f and g be functions, defined on subsets of $\mathbb{R}$, and suppose*

$$\lim_{x\to p} g(x) = q \text{ and } \lim_{v\to q} f(v) = L.$$

Let S be the domain of the composite $f\circ g$. Assume that:

(i) *p is a limit point of S;*

(ii) *p is exterior to the set $\{x : x\neq p,\, g(x)=q\}\cap S$. (In other words, p has a neighborhood U_0 with the property that there is no point in U_0, other than p itself, that is both in the domain of f and were g takes the value q.)*

Then $\lim_{x\to p} f\big(g(x)\big)$ exists and is equal to L:

$$\lim_{x\to p}(f\circ g)(x) = L.$$

Condition (ii) ensures that $g(x)$ avoids the 'forbidden' value q for all x that are used in determining $\lim_{x\to p} f(g(x))$. Notice that (ii) is automatically satisfied in the case $q = \infty$, for g, being a real valued function, never takes the value ∞. Another very convenient special case is when $g(x)$ *simply does not take the value* q, except possibly when $x = p$.

Proof. Let us begin by recalling what it means for $f(v) \to L$ as $v \to q$. Let W be any neighborhood of L. Then there is a neighborhood V of q such that

$$f(v) \in W \text{ for all } v \in V, \text{ with } v \neq q, \text{ that are in the domain of } f. \tag{7.15}$$

Next, since $g(x) \to q$ as $x \to p$, there is a neighborhood U of p such that

$$g(x) \in V \text{ for all } x \in U, \text{ with } x \neq p, \text{ that are in the domain of } g. \tag{7.16}$$

We will focus on the neighborhood U of p shrunk down by intersecting with U_0:

$$U_1 = U \cap U_0,$$

which is still a neighborhood of p, of course.

Now consider any point x in U_1, with $x \neq p$, for which $f(g(x))$ is defined and $g(x) \neq q$; we are given that such an x exists. By (7.16) we have $g(x) \in V$, and then by (7.15), we have $f\big(g(x)\big) \in W$.

Thus, starting with any neigborhood W of L we have produced a neighborhood U_1 of p such that $(f \circ g)(x) \in W$ for all $x \in U_1$ with $x \neq p$. $\boxed{\text{QED}}$

Chapter 8

Trigonometric Functions

We consider the trigonometric functions sin and cos. Though we don't really discuss completely precise mathematical definitions for these functions, we extract enough information about them from trigonometry to be able to do calculus with these functions. Eventually one can use the results of calculus to construct definitions for sin and cos that don't use geometry.

8.1 Measuring angles

What exactly is an angle? The most basic idea of an angle is that it is specified by two rays going out of a given vertex point.

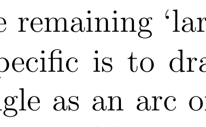
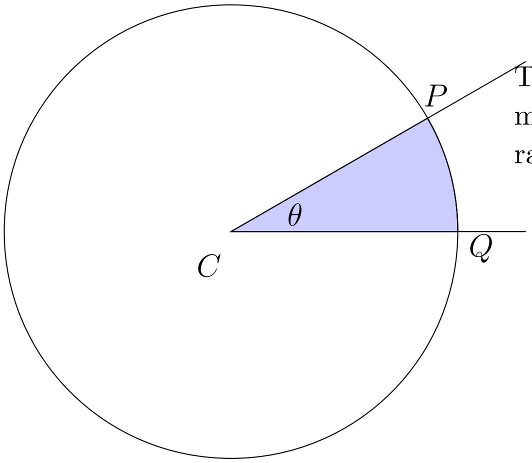


Figure 8.1: Angle as a pair of rays

This leaves open a small bit of ambiguity, as to whether we are thinking of the 'smaller' angle or the remaining 'larger' angle.

One way to be more specific is to draw a circle, with center C at the vertex, and think of the angle as an arc of the circle marked off by the two rays. To be more precise we could just think of a circle of radius 1, with



The area of the shaded sector gives a measure of the angle $\angle PCQ$ (taking the radius CP to be the unit of length).

Figure 8.2: Measuring angle using sectorial areas

center C at the vertex of the two rays, and then think of the angle as a sector marked off in the circle by the two rays.

Then the *radian measure* of the angle $\angle PCQ$ is taken to be *twice* the area of the sector PCQ.

Why twice? This is just to be consistent with historical practice and convention. Take, as an extreme example, the *full angle*, so that the sector PCQ is, in fact, the entire circular disk. The area of this disk is what is denoted

$$\pi$$

and so *the full circular angle* has radian measure 2π.

From this we can quickly see that 90^0, which specifies a quarter circle, has radian measure

$$\frac{1}{4}(2\pi) = \frac{\pi}{4}.$$

This discussion has one element of haziness: what do we mean by the area of a curved region? For this please turn back to the Introduction.

8.2 Geometric specification of sin, cos and tan

We will describe the geometric meaning of the measure of an angle and also that of $\sin\theta$ and $\cos\theta$.

Regardless of how an angle might be measured, the geometric meanings of sin, cos and tan of an acute angle are illustrated in the classical diagram shown in Figure 8.3. If the angle is specified by a pair of rays R_1 and R_2, initiating from a vertex C, we draw a circle, with center C, and take the radius to be the unit of length. The 'semichord' from R_2 to R_1 is the segment, perpendicular to R_1, that runs from the point Q where R_2 cuts the circle to a point on R_1. The length of the 'semi-chord' is the sin of the angle. The cos of the angle is the distance from the vertex C to the semi-chord. Then tan of the angle is the length of the segment tangent to the circle at Q to a point on R_1.

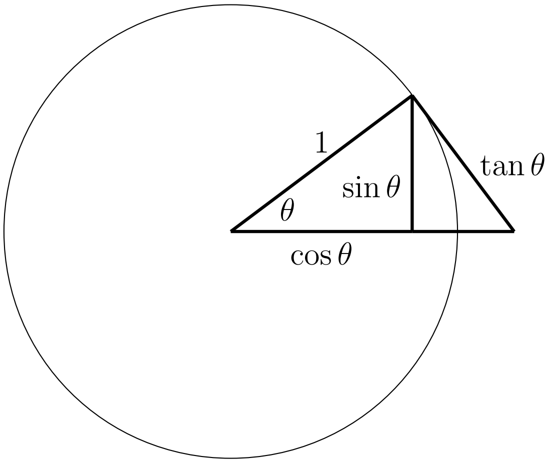


Figure 8.3: Classical definitions of sin, cos, and tan

The more cluttered Figure 8.4 provides more concrete formulas and also relates visually to the measurement of the angle θ in terms of the area of the sectorial region it cuts out of the circle.

The line through P perpendicular to CQ intersects the line CQ at a point B. Let

$$\begin{aligned} x &= CB \\ y &= QB. \end{aligned} \tag{8.1}$$

Here we take x to be *negative* if B is on the opposite side of C from P. We take y to be negative if $\theta > \pi$.

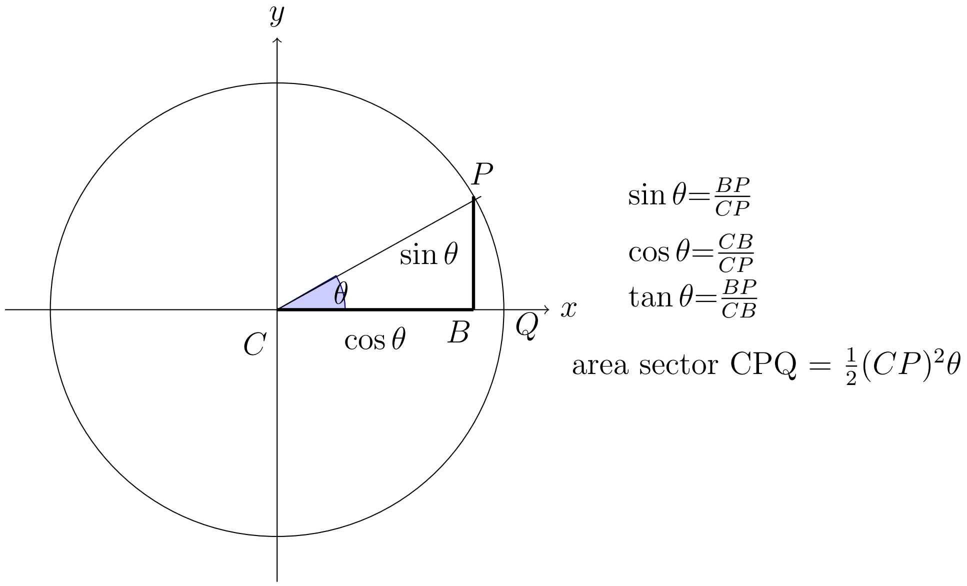

y
P
$\sin\theta$
θ
x
C
$\cos\theta$
B
Q
$\sin\theta = \frac{BP}{CP}$
$\cos\theta = \frac{CB}{CP}$
$\tan\theta = \frac{BP}{CB}$
area sector CPQ $= \frac{1}{2}(CP)^2\theta$

Figure 8.4: Measuring θ and $\sin\theta$, $\cos\theta$, and $\tan\theta$

The trigonometric functions are specified by:

$$\begin{aligned} \sin\theta &= \frac{QB}{CQ}; \\ \cos\theta &= \frac{CB}{CQ}; \\ \tan\theta &= \frac{QB}{CB}, \end{aligned} \tag{8.2}$$

where we leave $\tan\theta$ undefined if $\theta = \pi/2$ because in this case the denominator CB becomes 0.

From (8.2) it is also clear that

$$\tan\theta = \frac{\sin\theta}{\cos\theta} \tag{8.3}$$

as long as $\theta \neq \pi/2$.

Geometrically, to be consistent with the preceding discussion, it is sensible to define

$$\begin{aligned} \sin 0 &= 0; \\ \cos 0 &= 1; \\ \tan 0 &= 0. \end{aligned} \tag{8.4}$$

Observe also that

$$\begin{aligned}\sin \pi/2 &= 1;\\ \cos \pi/2 &= 0,\end{aligned} \tag{8.5}$$

and

$$\begin{aligned}\sin \pi &= 0;\\ \cos \pi &= -1;\\ \tan \pi &= 0.\end{aligned} \tag{8.6}$$

If the angle θ is increased to $\theta+2\pi$ then geometrically the point Q remains where it is. Thus we can define the trigonometric functions for values outside $[0, 2\pi)$ by requiring that

$$\begin{aligned}\sin(a + 2\pi) &= \sin a\\ \cos(a + 2\pi) &= \cos a;\\ \tan(a + 2\pi) &= \tan a,\end{aligned} \tag{8.7}$$

again as long as $\tan(a + 2\pi)$, and hence $\tan a$, is defined.

The above property of the trigonometric functions are summarized in words by saying that these functions are *periodic* and each has period 2π (the values repeat every time a is changed to $a + 2\pi$, and 2π is the least positive value with this property).

When $a \in (0, \pi)$, $\sin a$ is postive, and when $a \in (\pi, 2\pi)$ the value $\sin a$ is negative:

$$\sin a \begin{cases} > 0 & \text{if } a \in (0, \pi);\\ < 0 & \text{if } a \in (\pi, 2\pi).\end{cases} \tag{8.8}$$

For cos it is:

$$\cos a \begin{cases} > 0 & \text{if } a \in (-\pi/2, \pi/2);\\ < 0 & \text{if } a \in (\pi/2, 3\pi/2).\end{cases} \tag{8.9}$$

8.3 Reciprocals of sin, cos, and tan

The reciprocals of sin, cos and tan also have names:

$$\begin{aligned}\csc\theta &= \frac{1}{\sin\theta}\\ \sec\theta &= \frac{1}{\cos\theta}\\ \cot\theta &= \frac{1}{\tan\theta}\end{aligned} \tag{8.10}$$

whenever these reciprocals are meaningful (for instance, $\csc 0$ and $\sec(\pi/2)$ undefined).

8.4 Identities

If one angle of a right-angled triangle is θ then the other is $\pi/2-\theta$. This leads to the following identities:

$$\begin{aligned}\sin\left(\frac{\pi}{2}-\theta\right) &= \cos\theta\\ \cos\left(\frac{\pi}{2}-\theta\right) &= \sin\theta\\ \tan\left(\frac{\pi}{2}-\theta\right) &= \cot\theta.\end{aligned} \tag{8.11}$$

When an angle is replaced by its negative, it changes the sign of sin and tan but not of cos:

$$\begin{aligned}\sin(-a) &= -\sin a;\\ \cos(-a) &= \cos a;\\ \tan(-a) &= -\tan a,\end{aligned} \tag{8.12}$$

with the last holding if the tan values exist.

Pythagoras' theorem implies the enormously useful identity

$$\sin^2 a + \cos^2 a = 1, \tag{8.13}$$

for all $a \in \mathbb{R}$. Using this we can work out the value of sin, at least up to sign, from the value of cos:

$$\sin a = \pm\sqrt{1-\cos^2 a}. \tag{8.14}$$

The only way to decide whether it is + or whether it is − is to consider the value of a: if $a \in [0, \pi]$, or differs from such a value by an integer multiple of 2π, then $\sin a$ is ≥ 0.

Similar considerations hold for

$$\cos a = \pm\sqrt{1 - \sin^2 a}. \tag{8.15}$$

There are several relations among these reciprocal trigonometric functions that can be deduced from relations between sin, cos, and tan. For instance, dividing

$$\sin^2\theta + \cos^2\theta = 1$$

by $\cos^2\theta$ produces:

$$\frac{\sin^2\theta}{\cos^2\theta} + 1 = \frac{1}{\cos^2\theta},$$

which can be rewritten as

$$1 + \tan^2\theta = \sec^2\theta \tag{8.16}$$

for all θ for which $\tan\theta$ is defined. (For $\theta = \pm\pi/2$ one could define $\tan^2\theta$ as well as $\sec^2\theta$both to be ∞, and similarly for all the other trouble spots $\pm\pi/2$ plus integer multiples of 2π, and this would make (8.16) valid for all $\theta \in \mathbb{R}$.)

Very clever geometric arguments can be used to prove the trigonometric identities:

$$\begin{aligned}
\sin(a+b) &= \sin a\cos b + \sin b\cos a;\\
\cos(a+b) &= \cos a\cos b - \sin a\sin b;\\
\tan(a+b) &= \frac{\tan a + \tan b}{1 - \tan a\tan b},
\end{aligned} \tag{8.17}$$

where for the last identity we require, of course, that the tan values are actually defined.

There are some consequences of these addition formulas that are also useful. The simplest are obtained by taking $b = a$ in the preceding formulas.

Special cases of these are also very useful:

$$\begin{aligned}
\sin(2a) &= 2\sin a\cos b;\\
\cos(2a) &= \cos^2 a - \sin^2 a = 2\cos^2 a - 1 = 1 - 2\sin^2 a;\\
\tan(2a) &= \frac{2\tan a}{1 - \tan^2 a},
\end{aligned} \tag{8.18}$$

as long as the tan values are defined.

Now consider

$$\sin x - \sin y.$$

Suppose we choose a and b such that

$$\begin{aligned} x &= a + b \\ y &= a - b \end{aligned} \tag{8.19}$$

Then

$$\begin{aligned} \sin x - \sin y &= \sin(a + b) - \sin(a - b) \\ &= (\sin a \cos b + \sin b \cos a) - (\sin a \cos b - \sin b \cos a) \\ &= 2 \sin a \cos b. \end{aligned} \tag{8.20}$$

Now we need to substitute in the values of a and b in terms of x and y by solving (8.19). Adding the equations (8.19) gives

$$x + y = 2a,$$

and subtracting gives

$$x - y = (a + b) - (a - b) = 2b.$$

Thus

$$\begin{aligned} a &= \frac{1}{2}(x + y) \\ b &= \frac{1}{2}(x + y). \end{aligned} \tag{8.21}$$

Putting these into (8.20) produces

$$\sin x - \sin y = 2 \sin \frac{x - y}{2} \cos \frac{x + y}{2}. \tag{8.22}$$

Following the same line of reasoning for cos instead of sin gives us

$$\cos x - \cos y = -2 \sin \frac{x - y}{2} \sin \frac{x + y}{2}. \tag{8.23}$$

Before moving on, let us note that taking $-b$ in place of b in

$$\begin{aligned} \sin(a - b) &= \sin a \cos b - \sin b \cos a; \\ \cos(a - b) &= \cos a \cos b + \sin a \sin b; \\ \tan(a - b) &= \frac{\tan a - \tan b}{1 + \tan a \tan b}, \end{aligned} \tag{8.24}$$

8.5 Inequalities

The identity

$$\sin^2 a + \cos^2 a = 1$$

implies that neither $\sin a$ nor $\cos a$ can be bigger than 1 in magnitude:

$$\begin{aligned} |\sin a| &\leq 1; \\ |\cos a| &\leq 1. \end{aligned} \tag{8.25}$$

For example,

$$|5| = 5; \qquad |0| = 0; \quad \text{and } |-4| = 4.$$

Note, however, that both $\sin a$ and $\cos a$ do reach the values 1 and -1 repeatedly no matter how far away from 0 the value of a is:

$$\begin{aligned} \sin\left(\frac{\pi}{2} + 2\pi n\right) &= 1; \\ \sin\left(-\frac{\pi}{2} + 2\pi n\right) &= -1, \end{aligned} \tag{8.26}$$

for all integers $n \in \mathbb{Z}$. The same holds for cos:

$$\begin{aligned} \cos(2\pi n) &= 1; \\ \cos(\pi + 2\pi n) &= -1, \end{aligned} \tag{8.27}$$

for all integers $n \in \mathbb{Z}$.

Some geometric arguments with areas shows that

$$\cos x \leq \tfrac{\sin x}{x} \leq 1 \text{ for all } x \in (0, \pi/2]. \tag{8.28}$$

Since $\sin(-x)$ is $-\sin x$, we have

$$\frac{\sin(-x)}{-x} = \frac{\sin x}{x}.$$

Moreover, we also know that

$$\cos(-x) = \cos x.$$

Thus we can go over to the negative side as well:

$$\cos x \leq \tfrac{\sin x}{x} \leq 1 \text{ for all } x \in (-\pi/2, \pi/2) \text{ with } x \neq 0. \tag{8.29}$$

8.6 Limits for sin and cos

The functions sin and cos are continuous functions in the sense that their limits coincide with their values:

$$\begin{aligned} \lim_{x\to p} \sin x &= \sin p \\ \lim_{x\to p} \cos x &= \cos p, \end{aligned} \tag{8.30}$$

for all $p \in \mathbb{R}$. (We will return to the notion of continuous functions later.) In particular,

$$\lim_{x\to 0} \sin x = \sin 0 = 0$$

and

$$\lim_{x\to 0} \cos x = \cos 0 = 1.$$

We can explore the behavior of $\sin x$ for x near 0 more carefully. Recall the bounds for $(\sin x)/x$ near 0:

$$\cos x \leq \tfrac{\sin x}{x} \leq 1 \text{ for all } x \in (-\pi/2, \pi/2) \text{ with } x \neq 0. \tag{8.31}$$

We know that as $x \to 0$ we have $\cos x \to 1$. So $(\sin x)/x$ being between 1 and $\cos x$ also goes to the limit 1:

$$\lim_{x\to 0} \frac{\sin x}{x} = 1. \tag{8.32}$$

There are many useful and not-so-useful consequences of this limit. For instance,

$$\begin{aligned} \lim_{x\to 0} \frac{1-\cos x}{x^2} &= \lim_{x\to 0} \frac{(1-\cos x)(1+\cos x)}{x^2(1+\cos x)} \\ &= \lim_{x\to 0} \frac{1-\cos^2 x}{x^2(1+\cos x)} \\ &= \lim_{x\to 0} \frac{\sin^2 x}{x^2(1+\cos x)} \\ &= \lim_{x\to 0} \left(\frac{\sin x}{x}\right)^2 \frac{1}{1+\cos x} \\ &= 1^2 \cdot \frac{1}{1+1} \\ &= \frac{1}{2}. \end{aligned} \tag{8.33}$$

In summary

$$\lim_{x\to 0}\frac{1-\cos x}{x^2}=\frac{1}{2}. \tag{8.34}$$

Another very simply consequence of (8.32) is

$$\lim_{x\to 0}\frac{\sin Kx}{x}=K, \quad \text{for all } K\in\mathbb{R}. \tag{8.35}$$

This can be seen by observing that we can write

$$\frac{\sin Kx}{x}=K\frac{\sin Kx}{Kx}=K\frac{\sin y}{y},$$

where $y=Kx$. Clearly, y approaches 0 when $x\to 0$, and so we have

$$\lim_{x\to 0}\frac{\sin Kx}{x}=\lim_{y\to 0}K\frac{\sin y}{y}=K\cdot 1=K.$$

(To be honest, some reasoning is needed here to explain why one can pass from $x\to 0$ to $y\to 0$.)

8.7 Limits with $\sin(1/x)$

The functions $\sin(1/x)$ and $\cos(1/x)$ are badly behaved functions near the value $x=0$.

When $x=1/(\pi/2+2\pi n)$, for any integer n, we have $\sin(1/x)=1$:

$$\sin\left(\frac{1}{1/(\pi/2+2\pi n)}\right)=\sin\left(\pi/2+2\pi n\right)=1 \quad \text{for all } n\in\mathbb{Z}.$$

On the other hand if $x=3\pi/2+2\pi n$ then the value of sin is -1:

$$\sin\left(\frac{1}{1/(3\pi/2+2\pi n)}\right)=\sin\left(3\pi/2+2\pi n\right)=-1 \quad \text{for all } n\in\mathbb{Z}.$$

Now any neighborhood of 0 contains the values $1/(\pi/2+2\pi n)$ and $1/(3\pi/2+2\pi n)$ for large enough integers n. Thus, in every neighborhood of 0 there are values of x for which $\sin x$ is 1 and there are values of x for which the value of $\sin x$ is -1.

Thus

$$\inf_{x\in U, x\neq 0}\sin(1/x)=-1, \qquad \text{and} \qquad \sup_{x\in U, x\neq 0}\sin(1/x)=1 \tag{8.36}$$

for *every* neighborhood U of 0. Thus there cannot be a unique value lying between the sups and infs. Hence:

$$\lim_{x\to 0} \sin \tfrac{1}{x} \textit{ does not exist.} \tag{8.37}$$

The problem here is that $\sin 1/x$ fluctuates too much near $x = 0$. These fluctuations can be dampened out by multiplying by x; consider

$$f(x) = x \sin(1/x) \qquad \text{for } x \neq 0.$$

Since $\sin a$ is at most 1 in magnitude we have

$$-|x| \leq f(x) \leq |x| \quad \text{for all } x \neq 0.$$

If we let $x \to 0$ then clearly $|x| \to 0$, and so the squeeze theorem implies that $\lim_{x\to 0} f(x)$ exists and is 0:

$$\lim_{x\to 0} x \sin \frac{1}{x} = 0. \tag{8.38}$$

Notice that the 'product rule' does not work here:

$$\lim_{x\to 0} x \sin \frac{1}{x} = \left(\lim_{x\to 0} x\right)\left(\lim_{x\to 0} \sin \frac{1}{x}\right) \qquad \text{FAILS}$$

because on the right the limit $\lim_{x\to 0} \sin \frac{1}{x}$ does not exist.

8.8 Graphs of trigonometric functions

The graphs of sin and cos are waves, with sin passing through $(0, 0)$ and cos through $(1, 0)$.

The graph for sin is

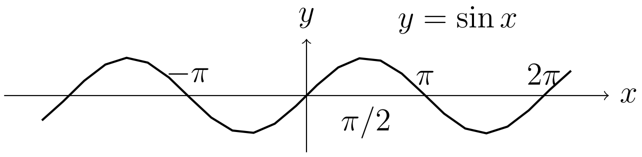


Figure 8.5: Graph of sin

The graph for tan blows up at $\pm\pi/2$, because

$$\lim_{x\to \pi/2+} \tan x = -\infty, \qquad \text{and} \qquad \lim_{x\to \pi/2-} \tan x = \infty,$$

and similarly at $-\pi/2$.

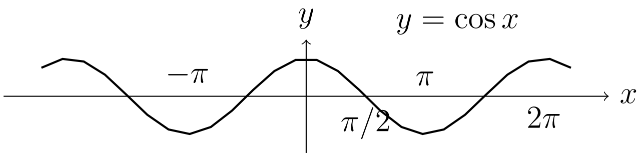


Figure 8.6: Graph of cos

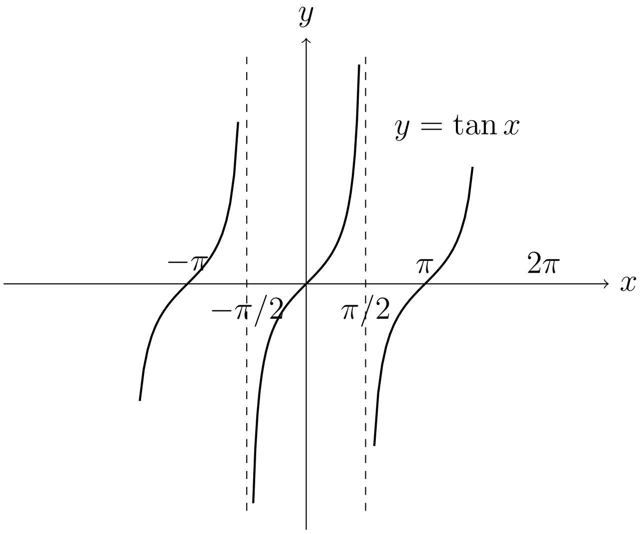


Figure 8.7: Graph of tan

8.9 Postcript on trigonometric functions

Once one has built up the full apparatus of calculus, with both derivatives and integrals, it is possible to reconstruct the functions sin, cos and tan directly in terms of calculus, without reference to any diagrams or traditional trigonometry. For example, here are the formulas for sin and cos that can be used to *define* them without using pictures:

$$\begin{aligned} \sin x &= x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \cdots \\ \cos x &= 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \cdots \end{aligned} \tag{8.39}$$

What the 'infinite sums' on the right mean exactly will be clear only after we have studied limits and sequences. It is, at this stage, impossible to see where the identities/definitions (8.39) come from.

Exercises on Limits

Write out the limits or explain as needed:

1. $\lim_{x\to 1} 5$
2. $\lim_{x\to 1}(x^2+4x-\frac{5}{x})$
3. $\lim_{x\to 1} \frac{x^2-9}{x-3}$
4. $\lim_{x\to 1} \frac{x^3-1}{x^2-1}$
5. $\lim_{x\to 1} \frac{x^4-1}{x^2-1}$
6. $\lim_{x\to\infty} \frac{1}{x^2}$
7. $\lim_{x\to\infty} \frac{4x^3-3x+2}{x^2-x+1}$
8. $\lim_{x\to\infty} \frac{5x^6-7x+2}{3x^6+x+2}$
9. $\lim_{x\to\infty} \frac{4x^3+\sin x}{2x^3+\sqrt{x}}$
10. $\lim_{x\to\infty} \frac{7x^5+x+\cos(x^3)}{2x^5-5x^2+1}$
11. $\lim_{x\to\infty} \left[\sqrt{x+1}-\sqrt{x}\right]$
12. $\lim_{x\to\infty} \left[\sqrt{3x^2+1}-\sqrt{x^2+1}\right]$
13. $\lim_{x\to\infty} \left[\sqrt{4x^4+2}-\sqrt{x^4+1}\right]$
14. $\lim_{x\to\infty} \frac{\sqrt{x+1}}{\sqrt{x}}$
15. $\lim_{x\to\infty} x\left[\sqrt{x^2+2}-\sqrt{x^2+1}\right]$
16. $\lim_{x\to\infty} \sqrt{x+2}\left[\sqrt{x+1}-\sqrt{x}\right]$
17. $\lim_{\theta\to 0} \frac{\sin(\theta^2)}{\theta^2}$
18. $\lim_{\theta\to 0} \frac{\sin^2(\theta)}{\theta^2}$
19. $\lim_{\theta\to\pi/6} \frac{\sin(\theta-\pi/6)}{\theta-\pi/6}$

20. $\lim_{x\to 0} x^2 1_{\mathbb{Q}}(x)$
21. $\lim_{x\to 0} x(1-x) 1_{\mathbb{Q}}(x)$
22. $\lim_{x\to 1} x(1-x) 1_{\mathbb{Q}}(x)$
23. Explain why $\lim_{x\to 3} x(x-1) 1_{\mathbb{Q}}(x)$ does not exist.
24. Explain why $\lim_{x\to\infty} \cos x$ does not exist.
25. Explain why $\lim_{x\to\infty} x \sin x$ does not exist.
26. Explain why $\lim_{x\to\infty} \frac{\sin x}{x} = 0$.
27. Explain why $\lim_{x\to\infty} \frac{\sin x}{\sqrt{x}} = 0$.

Chapter 9

Continuity

Continuous functions are functions that respect topological structure. They are also the easiest to work with in and therefore most suitable in applications.

9.1 Continuity at a point

A function f on a set $S \subset \mathbb{R}$ is said to be *continuous* at a point $p \in S$ if $f(x)$ approaches its actual value $f(p)$ when x approaches p:

$$\text{if } \lim_{x \to p} f(x) = f(p) \text{ we say } f \text{ is continuous at } p.$$

In case p is an isolated point of S we cannot work with $\lim_{x \to p} f(x)$, but surely there is no reason to view f as being not continuous at such a point. So we also say that f is continuous at p if p is an isolated point of S.

Here is a cleaner definition of continuity at p:

Definition 9.1.1 *A function f defined on a set $S \subset \mathbb{R}$ is said to be continuous at a point $p \in S$ if for every neighborhood W of $f(p)$ there is a neighborhood U of p such that*

$$f(x) \in W \text{ for all } x \in U.$$

9.2 Discontinuities

Sometimes a function is *discontinuous* (that is, not continuous) at a point p because the value $f(p)$ is, for whatever reason, not equal to $\lim_{x \to p} f(x)$ even though this limit exists. For example, for the function g given by

$$g(x) = \begin{cases} \frac{x^2-9}{x-3} & \text{if } x \neq 3; \\ 4 & \text{if } x = 3. \end{cases} \tag{9.1}$$

we have the limit

$$\lim_{x\to 3} g(x) = \lim_{x\to 3}(x+3) = 6,$$

which is not equal to the value $g(3)$. This type of discontiuity is *removable*, simple by changing the value of g at 3.

On the other hand there are more serious discontinuties. For example,

$$\lim_{x\to 0+} \frac{|x|}{x} = \lim_{x\to 0} \frac{x}{x} = 1,$$

whereas, approaching 0 from the left,

$$\lim_{x\to 0-} \frac{|x|}{x} = \lim_{x\to 0} \frac{-x}{x} = -1.$$

There is a *jump* from left to right, and there is no way to remove this discontinuity.

The function

$$f(x) = \begin{cases} \sin(1/x) & \text{for all } x \neq 0; \\ 1 & \text{if } x = 0 \end{cases}$$

has a more severe discontinuity at 0 because $\sin(1/x)$ doesn't even have a limit $x \to 0$.

9.3 Continuous functions

A function f is said to be *continuous* if f is continuous at every point where it is defined.

All polynomial functions, such as

$$5x^3 - 3x^4 + 7x^2 - 3x + 4$$

are continuous.

Here is a simple observation that is used often without mention:

Proposition 9.3.1 *If f is continuous at every point of a set S and if T is a nonempty subset of S then f is also continuous at every point of T.*

You can check this easily by consulting the definition of what it means to be continuous at a point.

If f is a function and T is a set contained inside the domain of definition of f then $f|T$, called the *restriction of f to T*, denotes the function whose domain is T and whose value at any $x \in T$ is $f(x)$.

For example, consider the function $1_{\mathbb{Q}}$ on $\mathbb{R}$ whose value is 1 on rationals and 0 on irrationals. The restriction

$$1_{\mathbb{Q}}|\mathbb{Q}$$

is the function defined on $\mathbb{Q}$ whose value at every point in $\mathbb{Q}$ is 1: in other words $1_{\mathbb{Q}}|\mathbb{Q}$ is just the constant function 1 on the set $\mathbb{Q}$.

The statement 'f is continuous on a set T' can have two different meanings:

(i) f is continuous at every point of T;

(ii) the restriction $f|T$ is continuous.

For example $1_{\mathbb{Q}}|\mathbb{Q}$ is certainly continuous but $1_{\mathbb{Q}}$ is not continuous at any point of $\mathbb{Q}$ (or at any point at all).

9.4 Two examples using $\mathbb{Q}$

The function

$$1_{\mathbb{Q}}$$

has the property that $\lim_{x\to p} 1_{\mathbb{Q}}(x)$ does not exist for *any* p. Hence this function is discontinuous *everywhere* on $\mathbb{R}$.

We can damp out the discontiuity at 0 as follows:

$$f(x) = x1_{\mathbb{Q}}(x) \quad \text{for all } x \in \mathbb{R}.$$

has the property that

$$\lim_{x\to 0} f(x) = 0 = f(0),$$

but $\lim_{x\to p} f(x)$ does not exist for any $p \neq 0$. Hence f is continuous *at exactly one point*, that being 0.

To produce a function continuous at only the points 1 and 5 we take $1_{\mathbb{Q}}(x)$ and multiply it with a function that is 0 at exactly 1 and 5; for example,

$$x \mapsto (x-1)(x-5)1_{\mathbb{Q}}(x)$$

is continuous at 1 and at 5 but nowhere else.

Can you manufacture a function that is continuous at exactly a given set of points and nowhere else? Is there a function that continuous at every point of $(0,1)$ but at no other point?

9.5 Composites of continuous functions

Proposition 9.5.1 *Suppose f and g are functions defined on subsets of $\mathbb{R}$, and suppose $f \circ g$ is defined on a neighborhood of some $p \in \mathbb{R}$. If g is continuous at p and f is continuous at $g(p)$ then $f \circ g$ is continuous at p.*

Proof. Let W be a neighborhood of $L = f(q)$, where $q = g(p)$. Then, by continuity of f at q, there is a neighborhood V of q such that $f(v) \in W$ for all $v \in V$ in the domain of f. Next, by continuity of g at p, there is a neighborhood U of p such that $g(x) \in V$ for all $x \in U$ in the domain of g. Hence if $x \in U$ is in the domain of $f \circ g$ then $f\big(g(x)\big) \in W$. This proves that $f \circ g$ is continuous at p. $\boxed{\text{QED}}$

9.6 Continuity on $\mathbb{R}^*$

In calculus we work with functions defined on subsets of $\mathbb{R}$ and having values in $\mathbb{R}$. However, occasionally, it is useful to allow infinite values as well. No great additional work is needed for this; the definition remains exactly as before:

a function $F : S \to \mathbb{R}^*$ is continuous at $p \in S$ if either p is an isolated point of S or if $\lim_{x\to p} F(x) = F(p)$.

An equivalent alternative form is again just as before: *F is continuous at p if for any neighborhood W of $F(p)$ there is a neighborhood U of p such that $F(x) \in W$ for all $x \in U$.*

If $f : (a,b) \to \mathbb{R}$ is continuous and $\lim_{x\to a} f(x)$ exists then we can extend f to a continuous function $F : [a,b) \to \mathbb{R}^*$ by setting

$$F(a) = \lim_{x\to a} f(x).$$

Of course, the same applies for the other endpoint b.

Chapter 10

The Intermediate Value Theorem

Consider the function f given by

$$f(x) = x^3 - x^2 - 2x + 1 \qquad \text{for all } x \in \mathbb{R}.$$

Figure 10.1 is a sketch of its graph for $x \in [-1.5, 2]$. We can check easily that

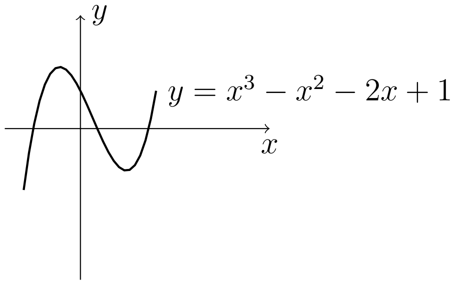


Figure 10.1: Graph of $x^3 - x^2 - 2x + 1$

$$f(-1) = 1 \qquad \text{and} \qquad f(1) = -1.$$

Intuitively it is clear, from the continuous nature of the graph of $y = f(x)$, that there must be a point $p \in [-1, 1]$ where $f(p)$ is 0. The *intermediate value theorem*, which we study in this chapter, guarantees the existence of such a point; thus, from this theorem it follows that there is a solution of the equation

$$x^3 - x^2 - 2x + 1 = 0$$

on the interval $-1, 1]$.

Its essence remains valid in settings far beyond the real line, but even this first glimpse of the idea, on $\mathbb{R}$, is of great use.

10.1 Inequalities from limits

Suppose we know that a function f has the limit

$$\lim_{x\to 5} f(x) = 9.$$

Then $f(x)$ is close to 9, say within a distance of 1, if x is close enough to 5 (but not 5); thus there must be a neighborhood U of 5 for which

$$8 < f(x) \qquad \text{and} \qquad f(x) < 10,$$

for all $x \in U$ with $x \neq 5$. We summarize this idea in:

Proposition 10.1.1 *Let f be a function on a set S, and $p \in \mathbb{R}^*$ a point for which $\lim_{x\to p} f(x)$ exists. If K is less than this limit then there is a neighborhood U on which$K < f(x)$ for all $x \in U$, $x \neq p$; thus,*

$$\textit{if } K < \lim_{x\to p} f(x) \textit{ then } K < f(x) \textit{ for all } x \in U,\ x \neq p \tag{10.1}$$

for some neighborhood U of p. Similarly,

$$\textit{if } M > \lim_{x\to p} f(x) \textit{ then } M > f(x) \textit{ for all } x \in U,\ x \neq p \tag{10.2}$$

for some neighborhood U of p.

The proof proceeds a little bit differently from the intuition if we use our sup-inf definition of limit:
Proof. Let

$$L = \lim_{x\to p} f(x).$$

This means that L is the *unique* value satisfying

$$\inf_{x\in U, x\neq p} f(x) \leq L \leq \sup_{x\in U, x\neq p} f(x)$$

for all neighborhoods U of p. So if $K < L$ then K does not lie between all such infs and sups; thus, there is some neighborhood U of p such that K

does not lie between $\inf_{x\in U, x\neq p} f(x)$ and $\sup_{x\in U, x\neq p} f(x)$. Since $K < L$, the only possibility left is

$$K < \inf_{x\in U, x\neq p} f(x).$$

This proves (10.1).

The result (10.2) for $M > \lim_{x\to p} f(x)$ follows by a similar argument. $\boxed{\text{QED}}$

10.2 Intermediate Value Theorem

The completeness property of the real line has one big consequence for continuous functions: if f is continuous on the interval $[a, b]$ then $f(x)$ runs through all the values between $f(a)$ and $f(b)$ as x runs over $[a, b]$:

Theorem 10.2.1 *Let f be a continuous function on $[a, b]$, where $a, b \in \mathbb{R}$ with $a < b$. Let t be any real number between $f(a)$ and $f(b)$:*

$$f(a) \leq t \leq f(b) \quad \textit{or} \quad f(b) \leq t \leq f(a).$$

Then there is a point $s \in [a, b]$ for which $f(s) = t$.

Proof. If t happens to be equal to $f(a)$ then we are done; just take $s = a$. Similarly if $t = f(b)$.

Suppose then that t is neither $f(a)$ nor $f(b)$, and so lies *strictly between* them. If $f(a) < f(b)$ this means that $f(a) < t < f(b)$, whereas if $f(a) > f(b)$ then $f(b) < t < f(a)$.

Suppose

$$f(a) < t < f(b).$$

Let S be the set of all $x \in [a, b]$ for which $f(x) < t$:

$$S = \{x \in [a, b] \,:\, f(x) < t\}.$$

For instance, $a \in S$. Moreover, b is an upper bound for S, because S is inside $[a, b]$. In fact $a < s < b$, because of Proposition 10.1.1.

Then by the completeness property for $\mathbb{R}$ there is a least upper bound $s = \sup S$, and this, of course, also lies in $[a, b]$. We claim that $f(s)$ equals t. Consider any neighborhood U of s of the form

$$U = (s - \delta, s + \delta),$$

where δ is any positive real number. Since s is an upper bound of S, any point p of S strictly to the right of s (that is, $p > s$) is not in S, and so

$$f(p) > t,$$

for such $p \in [a, b]$. Then

$$\sup_{x \in U, x \neq s} f(x) > t$$

Since s is the *least* upper bound of S, any point $p \in U$ for which $p < s$ is *not* an upper bound of S and so there is some $q \in S$ with $q > p$. Of course $q \leq s$, since s is an upper bound of S. Hence q, lying between p and s, is in the neighborhood U. Since $q \in S$ we have

$$f(q) < t.$$

This shows that the inf of f over U, even excluding the point s, is $< t$:

$$\inf_{x \in U, x \neq s} f(x) < t.$$

Thus t satisfies:

$$\inf_{x \in U, x \neq p} f(x) < t < \sup_{x \in U, x \neq s} f(x)$$

for every neighborhood U of p. Since f is given to be continuous at s we know that

$$f(s) = \lim_{x \to s} f(x).$$

Hence t must be $f(s)$. $\boxed{\text{QED}}$

10.3 Intermediate Value Theorem: a second formulation

Here is another formulation of the intermediate value theorem:

Theorem 10.3.1 *If f is continuous on an interval J then the image*

$$f(J) \stackrel{\text{def}}{=} \{f(x) \,:\, x \in J\}$$

is also an interval.

Proof. To prove that $f(J)$ is an interval we need only to show that all the numbers between any two distinct values $y_1, y_2 \in f(J)$ also lie in $f(J)$. Thus consider a point t satisfying

$$y_1 < t < y_2.$$

Since $y_1 \in f(J)$ we have

$$y_1 = f(a),$$

for some $a \in J$, and since $y_2 \in J$ then

$$y_2 = f(b)$$

for some $b \in J$. Thus, f is continuous on $[a, b]$ and t lies between $f(a)$ and $f(b)$. Then by Theorem 10.2.1, there is a point $s \in [a, b]$ for which

$$f(s) = t.$$

Since $s \in [a, b]$ and a and b are points of the *interval* J it follows that s also lies in J. Thus any point t between y_1 and y_2 is of the form $t = f(s)$, with $s \in J$, which just means that $t \in J$. Thus, J is indeed an interval. $\boxed{\text{QED}}$

10.4 Intermediate Value Theorem: an application

The number

$$7^{3/4}$$

is the positive real number whose 4-th power is $(10)^3$:

$$\left(7^{3/4}\right)^4 = 7^3.$$

But how do we know that such a real number exists? We can obtain existence by using the intermediate value theorem.

Consider the function

$$q(x) = x^4 \quad \text{for all } x \in \mathbb{R}.$$

This is clearly continuous, and from

$$q(0) = 0 \quad \text{and} \quad q(7) = 7^4$$

we see that the number 7^3 lies between these extremes:

$$q(0) < 7^3 < q(7).$$

Hence, by the intermediate value theorem, there is a real number $s \in (0, 7)$ for which

$$s^4 = 7^3.$$

Could there be another positive real number s_* whose 4-th power is also 7^3? The answer is no, because if $s > s_* > 0$ then s_*^4 is $< s^4 = 7^3$, whereas if $s < s_*$ then $s_*^4 > s^4 = 7^3$. Thus there is a *unique positive real number* whose 4-th power is 7^3. This number is denoted

$$7^{3/4}.$$

In this was one can see that

$$x^y$$

exists for all positive real x and all rational y.

Returning to $7^{3/4}$ we can extract some more information: we saw that s actually lies between 0 and 7. But we can sharpen this much further. Since $7^3 = 343$ we have

$$q(4) = 4^4 = 256 < 7^3 < q(5) = 5^4 = 625.$$

Hence $7^{3/4}$ actually lies between 4 and 5. With more work we can narrow down the location of $7^{3/4}$ systematically.

It is clear that not much is special about the numbers 7 and 3/4 in this discussion. The intermediate value theorem (and, more fundamentally, the completeness of $\mathbb{R}$) shows that for any real number $x \geq 0$ and any rational number $r = p/q$, with $p, q \in \mathbb{Z}$ and $q \neq 0$, there is a unique non-negative real number x^r which satisfies

$$(x^{p/q})^q = x^p.$$

10.5 Locating roots

Consider the equation

$$x^7 - 3x + 1 = 0.$$

There is no systematic way to work out exact solutions of equations such as this. However, there are many ways of determining information about the solutions as well as finding very good approximations to them. Here let us see how the intermediate value theorem shows that there are solutions of the equation and helps localize them somewhat.

Consider the function

$$f(x) = x^7 - 3x - 1 \qquad \text{for all } x \in \mathbb{R}.$$

This is clearly continuous. Let us check a few values of f:

$$f(-2) = -123, \quad f(-1) = 1, \qquad f(0) = -1, \qquad f(1) = -3, \qquad f(2) = 121.$$

Since f is continuous and 0 lies between $f(-2)$ and $f(-1)$:

$$f(-2) = -123 \,<\, 0 \,<\, 1 = f(-1),$$

it follows by the intermediate value theorem that

there is a point $p \in (-2, -1)$ *where* $f(p) = 0$.

This means that the equation

$$x^7 - 3x + 1 = 0 \tag{10.3}$$

has a solution on the interval $(-2, -1)$.

By the same reasoning we see that the equation (10.3) also has a solution in $(-1, 0)$ and and a solution lying in $(1, 2)$.

One way of pinning down the location of the solutions of (10.3) would be to divide, say the interval $(1, 2)$ into ten pieces, each of width .1, and checking the values of f at the points

$$f(1), f(1.1), f(1.2), \ldots, f(1.9), f(2),$$

to see where f changes from negative to positive. We can calculate

$$f(1.1) \simeq -0.35 \qquad f(1.2) \simeq 0.98$$

and this tells us there is a root (which means the same as 'solution') of the equation (10.3) in the interval $(1.1, 1.2)$. Next, repeating the same strategy by dividing $(1.1, 1.2)$ into ten pieces and calculating the values

$$f(1.11),\ f(1.12),\ f(1.13),\ f(1.14),\ f(1.15),\ \ldots, f(1.19),\ f(1.2),$$

we observe that

$$f(1.13) \simeq -0.037, \qquad \text{and} \qquad f(1.14) \simeq 0.082.$$

Thus, there is a root in the interval

$$(1.13, 1.14).$$

This is a slow and inefficient process, but a first process nonetheless to systematically pin down a root of an equation. Later, with the use of calculus, we can study much faster methods for locating roots.

Chapter 11

Inverse Functions

In this chapter we show by using the intermediate value theorem that equations of the form

$$y = f(x)$$

can be 'solved' for a good class of continuous functions f. The solution is then displayed as

$$x = f^{-1}(y),$$

and f^{-1} is the called the *inverse* of the function f.

11.1 Inverse trigonometric functions

The graph of $y = \sin x$ oscillates between -1 and 1.

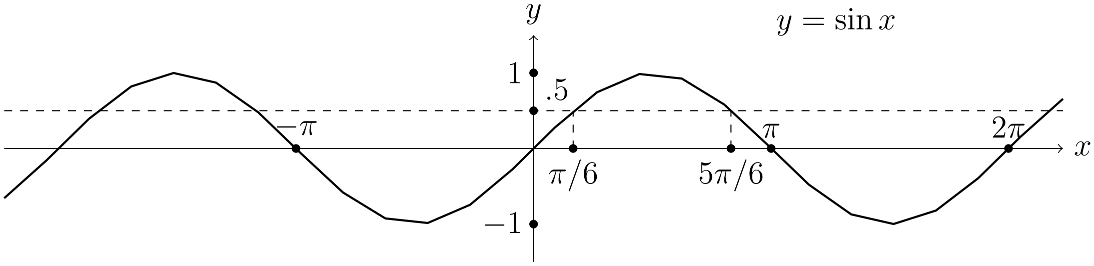


Figure 11.1: Graph of sin.

If we try to solve an equation such as

$$\sin \phi = .5$$

there are infinitely many values for ϕ:

$$\frac{\pi}{6},\ \frac{5\pi}{6},\ \frac{\pi}{6}+2\pi,\ \frac{5\pi}{6}+2\pi,\ \frac{\pi}{6}-2\pi,\ \frac{5\pi}{6}-2\pi,\ \frac{\pi}{6}+4\pi,\ \frac{5\pi}{6}+4\pi,\ \ldots$$

Each of these could be thought of as an 'inverse sin' for the value .5 in the sense that the sin of each of these is .5. However, to avoid ambiguity we can focus on just the value $\pi/6$: what makes it unique is that it is the only value between $-\pi/2$ and $\pi/2$ whose sin is .5.

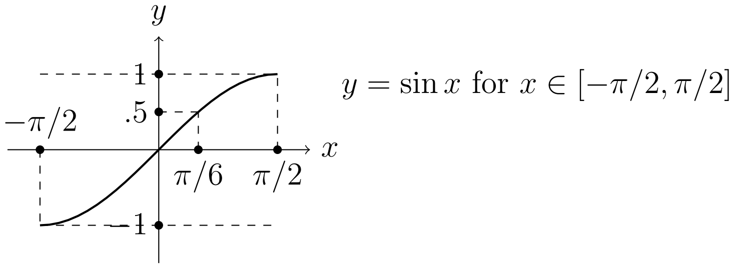


Figure 11.2: Graph of sin over $[-\pi/2, \pi/2]$.

We define arcsin(.5) to be $\pi/6$:

$$\arcsin(.5) \overset{\text{def}}{=} \sin^{-1}(.5) \overset{\text{def}}{=} \pi/6.$$

More generally

$$\sin^{-1}(w) \text{ is the unique value in } [-\pi/2, \pi/2] \text{ whose sin is } w, \tag{11.1}$$

that is,

$$\sin\left(\sin^{-1} w\right) = w \qquad \text{and } \sin^{-1} w \in [-\pi/2, \pi/2]. \tag{11.2}$$

Thus,

$$y = \sin^{-1} x \text{ means that } y \in [-\pi/2, \pi/2] \text{ and } \sin y \text{ is } x. \tag{11.3}$$

Since the values sin always lie between -1 and 1, there is no value whose sin is 2; thus

$$\sin^{-1} x \text{ is not defined, as a real number, if } x \text{ is not in } [-1, 1].$$

On the positive side,

Proposition 11.1.1 *If* $A \in [-1,1]$ *then there exists a unique* $B \in [-\pi/2, \pi/2]$ *for which* $\sin B = A$.

Proof. The function sin is continuous on $[-\pi/2, \pi/2]$ and the end point values are

$$\sin(-\pi/1) = -1, \quad \text{and} \quad \sin(\pi/2) = 1.$$

Therefore, by the intermediate value theorem, for any $A \in [-1,1]$ there is a $B \in [-\pi/2, \pi/2]$ for which

$$A = \sin B.$$

To see that B is unique simply observe that the function $y = \sin x$ is *strictly* increasing on $[-\pi/2, \pi/2]$ and two different values of x could not have the same value for $y = \sin x$. $\boxed{\text{QED}}$

Thus,

$$\sin^{-1} x \text{ is defined for all } x \in [-1,]1].$$

We can run through the same arguments, with minor changes, for the function cos. We have to be careful to observe that $y = \cos x$ is *not* strictly increasing on $[-\pi/2, \pi/2]$. For example both $\pi/3$ and $-\pi/3$ have cos equal to .5:

$$\cos(-\pi/3) = \cos(\pi/3) = .5.$$

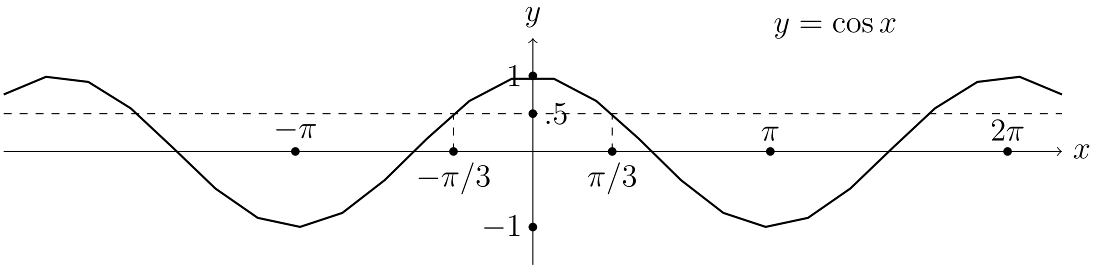


Figure 11.3: Graph of cos.

One possibility is to work with the interval $[-\pi, 0]$ on which cos is strictly increasing, but there is a bias against using negative values when positives would work. So, instead we use the interval

$$[0, \pi]$$

on which cos is strictly *decreasing*.

Running through the argument we conclude that

for every $A \in [-1,1]$ there is a unique $B \in [0,\pi]$ for which $\cos B = A$.

The unique value y in $[0,\pi]$ for which $\cos y$ is x is denoted $\cos^{-1} x$:

$$\cos^{-1}(x) \text{ is the unique value in } [0,\pi] \text{ whose cos is } x, \tag{11.4}$$

or, equivalently,

$$y = \cos^{-1} x \text{ means that } y \in [0,\pi] \text{ and } \cos y \text{ is } x. \tag{11.5}$$

We can run the same reasoning for tan as well and see that

for every $A \in \mathbb{R}$ there is a unique $B \in [-\pi/2, \pi/2]$ for which $\tan B = A$.

$$\tan^{-1}(x) \text{ is the unique value in } [-\pi/2,\pi/2] \text{ whose tan is } x, \tag{11.6}$$

or, equivalently,

$$y = \tan^{-1} x \text{ means that } y \in [-\pi/2,\pi/2] \text{ and } \tan y \text{ is } x. \tag{11.7}$$

11.2 Monotone functions: terminology

We say that a function f is *increasing* if

$$f(s) \leq f(t)$$

for all s, t in the domain of f for which $s \leq t$. We say that f is *strictly increasing* if

$$f(s) < f(t)$$

for all s, t in the domain of f for which $s < t$.

A function f is *decreasing* if

$$f(s) \geq f(t)$$

for all s, t in the domain of f for which $s \leq t$. We say that f is *strictly decreasing* if

$$f(s) > f(t)$$

for all s, t in the domain of f for which $s < t$.

Clearly, a function f is (strictly) increasing if and only if $-f$ is (strictly) decreasing. For this reason we will often state or prove results just for increasing functions, it being generally understood from context that the corresponding result for decreasing functions also holds.

A *monotone* function is a function that is increasing or that is decreasing. We say f is *strictly monotone* if f is stricly increasing or strictly decreasing.

The function

$$x_+ = \max\{x, 0\} \tag{11.8}$$

is an increasing function that is not strictly increasing.

11.3 Inverse functions

Here is a somewhat strange result, guaranteeing continuity of certain types of strictly monotone functions:

Proposition 11.3.1 *If g is a strictly monotone function defined on a set $S \subset \mathbb{R}$ such that the range $g(S)$ is an interval then g is continuous.*

The ideas developed for arcsin and arccos can be summarized in a general way:

Proposition 11.3.2 *If f is a continuous strictly monotone function on an interval $U \subset \mathbb{R}$ then the range V of f is also an interval, and there is a unique function f^{-1} defined on V such that*

$$f^{-1}\big(f(x)\big) = x \qquad \textit{for all } x \in U. \tag{11.9}$$

This inverse function f^{-1} *is continuous.*

The inverse function also satisfies

$$f\big(f^{-1}(y)\big) = y \qquad \textit{for all } y \in V. \tag{11.10}$$

For example, the inverse of the function

$$[0, \infty) \to [0, \infty) : x \mapsto x^2$$

is the function

$$[0, \infty) \to [0, \infty) : A \mapsto \sqrt{A}.$$

Proof. We work with the case when f is strictly increasing; the case of strictly decreasing is settled in an exactly similar way (or by applying the result for strictly increasing functions to $-f$ in place of f).

To show that V is an interval we have to show that if $c, d \in V$, with $c < d$, then every point between c and d is also in the range V of f. For $c, d \in V$ we have

$$c = f(a) \qquad \text{and} \qquad d = f(b),$$

for some $a, b \in U$. Now consider a point q strictly between c and d:

$$c < q < d.$$

This means

$$f(a) < q < f(b).$$

By the intermediate value theorem (keep in mind f is continuous there is a point $p \in (a, b)$ for which

$$q = f(p).$$

This means $q \in V$, the range of V. Thus V is an interval.

Now defined f^{-1} on V as follows. If $y \in V$ then there is a point $x \in U$ with $f(x) = y$. There cannot be any other point in U whose image under f is also y, because f is strictly increasing (points strictly below/above x are mapped by f to points strictly below/above y). Set

$$f^{-1}(y) = x \qquad \text{if } f(x) = y.$$

This proves (11.9), by simply writing in the value $f(x)$ in place of y in $f^{-1}(y) = x$.

Applying f to both sides of $f^{-1}(y) = x$ shows that

$$f\big(f^{-1}(y)\big) = f(x) = y,$$

which proves (11.10).

Continuity of f^{-1} follows on applying Proposition 11.3.1 to the function $g = f^{-1}$, whose range is the interval U. $\boxed{\text{QED}}$

Suppose f is a strictly increasing function defined on a set S. Then $f(S)$ can be thought of as S with the points p renamed as $f(p)$, and the ordering of the points is preserved:

$$f(s) < f(t) \quad \text{if and only if} \quad s < t.$$

Thus, $f(S)$ contains a largest element if and only if S contains a largest element, and $f(S)$ contains a smallest element if and only if S contains a smallest element. This gives us:

Proposition 11.3.3 *If f is a continuous, strictly increasing function, then for any interval J in the domain of f, the image $f(J)$ is of the same type as J; specifically,*

(i) *if $J = [a, b]$ then $f(J) = [f(a), f(b)]$;*

(ii) *if $J = (a, b]$ then $f(J) = (c, d]$, where $c = \inf f(J)$ and $d = \sup f(J) = f(b)$;*

(iii) *if $J = [a, b)$ then $f(J) = [c, d)$, where $c = \inf f(J) = f(a)$ and $d = \sup f(J)$;*

(iv) *if $J = (a, b)$ then $f(J) = (c, d)$, where $c = \inf f(J)$ and $d = \sup f(J)$.*

Chapter 12

Maxima and Minima

A fundamental feature of continuous functions is that they *attain* maximum and minimum values on certain types of sets such as closed intervals $[a, b]$, for $a, b \in \mathbb{R}$ with $a < b$.

12.1 Maxima and Minima

The completeness property of the real line has another big consequence for continuous functions: if f is continuous on the interval $[a, b]$ then $f(x)$ actually attains a maximum value at some point and a minimum value on the interval $[a, b]$.

Theorem 12.1.1 *Let f be a continuous function on $[a, b]$, where $a, b \in \mathbb{R}$ with $a < b$. Then there exist $c, d \in [a, b]$ such that*

$$\begin{aligned} f(c) &= \inf_{x \in [a,b]} f(x) \\ f(d) &= \sup_{x \in [a,b]} f(x). \end{aligned} \tag{12.1}$$

Before proceeding to logical reasoning here is our strategy for finding a point where f reaches the value

$$M = \sup_{x \in [a,b]} f(x).$$

Let us follow a point t, starting at a and moving to the right towards b and keep track of the 'running supremum'

$$S_f(t) = \sup_{x \in [a,t]} f(x).$$

If $f(a)$ itself is already the maximum M then we are done; assuming then that $f(a) < M$, surely the 'first exit time' t when $S_f(t)$ escapes from below the value M is where f actually takes the value M. Thus our guess for d is

$$d_* \stackrel{\text{def}}{=} \sup B_M, \tag{12.2}$$

where B_M is thet set of all t for which $S_f(t)$ is below M:

$$B_M = \{t \in [a, b] \, : S_f(t) < M\}\,. \tag{12.3}$$

(We are assuming the initial value $f(a)$ isn't already M.) It is useful to have in mind a graph of S_f: it increases (possibly remaining constant on stretches of values of t) and once it hits the value M it stays there all the way to $t = b$.

It is intuitively clear that for t to the left of d_* the value $S_f(t)$ is $< M$ whereas for any t to the right of d_* the value of $S_f(t)$ is M; this would imply that that the supremum of f on any neighborhood of d_* is in fact M. Then from the definition of limit $\lim_{x\to d_*} f(x)$ as the *unique* value between suprema and infima of f over neighborhoods of d_* we would then have $\lim_{x\to d_*} f(x) = M$; continuity of f at d_* would then imply $f(d_*) = M$.

Observe that if $S_f(x) < M$ then, since d_* is an upper bound of B_M, it follows that $d_* \geq x$. Thus, *no point to the right of d_* has S_f-value $< M$.* Hence:

$$S_f(x) = M \qquad \text{for all } x > d_*. \tag{12.4}$$

Proof of Theorem 12.1.1. We show only the existence of a point d where f attains its maximum value. The argument for minimum is exactly similar (or we can use the trick of applying the maximum result to $-f$ in place of f to find where f is minimum.)

Let us go through the remaining argument slowly, breaking it up into pieces.

If $f(a)$ happens to be equal to M then, of course, we are done, on taking d to be a. Suppose from now on that $f(a) < M$. This implies, in particular, that the set B_M is not empty, containing at least the point a.

Consider now any neighborhood U of d_*. Choose any $r \in U \cap [a, b]$ with $r > d_*$; what if $d_* = b$? We will deal with that case later, assuming for now that $d_* < b$. Then, as already noted before in (12.4), $S_f(r) = M$. Consequently,

$$\sup_{x\in U\cap[a,b]} f(x) = M.$$

Hence M satisfies

$$\inf_{x\in U\cap[a,b]} f(x) \leq M \leq \sup_{x\in U\cap[a,b]} f(x),$$

wth the second $\leq$ being actually an equality. This is true for *any* neighborhood U of d_*. Therefore, by our definition of limit,

$$\lim_{x\to d_*} f(x) = M.$$

But f is continuous at d_*. Hence

$$f(d_*) = M,$$

and we are done.

Lastly suppose $d_* = b$. Then taking any $q \in [a,b]$ with$q < b$, we know that q is *not* an upper bound of B_M (for $d_* = b$ is the *least* upper bound of B_M). So there is a $p > q$ in $[a,b]$ which is in B_M, and this means $\sup_{x\in[a,p]} f(x) < M$. Therefore also

$$\sup_{x\in[a,q]} f(x) < M.$$

But since $\sup_{x\in[a,b]} f(x)$ is M we must have $\sup_{x\in(q,b]} f(x) = M$. Thus the supremum of f over every neighborhood of d_* (which is b) is M. Then by the arhument used in the previous paragraph it follows again that $f(d_*) = M$.

The result for $\inf_{x\in[a,b]} f(x)$ is obtained similarly or just applying the result for sup to the function $-f$ instead of f. $\boxed{\text{QED}}$

The preceding heavily used result works for functions defined on closed intervals $[a,b]$, with $a, b \in \mathbb{R}$. But what of functions defined on other types of intervals? For example, for the function

$$\frac{1}{x} \qquad \text{for } x \in (0,\infty)$$

it is clear that the function is trying to reach its supremum ∞ at the left endpoint 0 and its infimum 0 at the right endpoint ∞. Figure 12.1 shows the graph of the function given on $(0,\infty)$ by $x^2 + \frac{2}{x} - 2$. The function has sup equal to ∞, which is the value it is trying to reach at both endpoints 0 and ∞ of the interval $(0,\infty)$; the inf occurs at $x = 1$ and the corresponding minimum value is $1^2 + \frac{2}{1} - 2 = 1$.

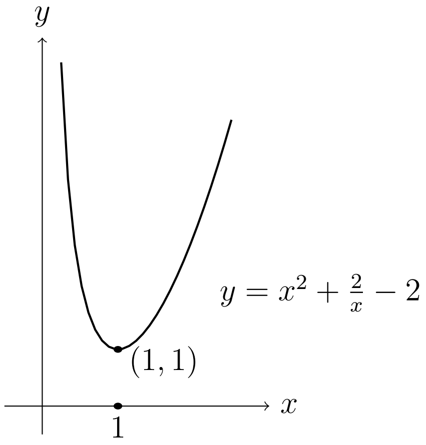


Figure 12.1: Graph of $x^2 + \frac{2}{x} - 2$, for $x > 0$.

Proposition 12.1.1 *Suppose f is a continuous function on an interval $U \subset \mathbb{R}$, with $a, b \in \mathbb{R}^*$ being the left and right endpoints, and suppose that both $L_a = \lim_{x \to a} f(x)$ and $L_b = \lim_{x \to b} f(x)$ exist and are in $\mathbb{R}$ (finite).*

Then either f attains a maximum value in the interior of U or $\sup_{x \in U} f(x)$ is the larger of the endpoint limits L_a and L_b.

Moreover, either f attains its minimum value at some point in the interior of U or $\inf_{x \in U} f(x)$ is the least of the two endpoint limits L_a and L_b.

We will not work through the proof but sketch the ideas. Observe that we can extend the function f to be defined at the endpoints a and b (if one or both of them is not already in U) but setting $f(a) = L_a$ and $f(b) = L_b$. Then f is defined on the interval $[a, b] \subset \mathbb{R}^*$ (denoting the left endpoint of U by a), and f is allowed to take the values $\pm\infty$ at the endpoints a and b. The resut of the argument follows the proof of Theorem 12.1.1.

12.2 Maxima/minima with infinities

The arguments used to prove existence of maxima and minima work without much change for functions defined on subsets of $\mathbb{R}^*$ and with values in $\mathbb{R}^*$:

Proposition 12.2.1 *If $F : [a, b] \to \mathbb{R}^*$ is continuous function, where $a, b \in \mathbb{R}^*$ with $a \leq b$, then F attains a maximum value and a minimum value on*

$[a, b]$. *Thus, there exist points* $c, d \in [a, b]$ *such that*

$$\begin{aligned} F(c) &= \inf_{x \in [a,b]} F(x) \\ F(d) &= \sup_{x \in [a,b]} F(x). \end{aligned} \tag{12.5}$$

12.3 Closed and bounded sets

Consider a set $K \subset \mathbb{R}$ that is closed and that lies inside an interval $[a, b]$, where $a, b \in \mathbb{R}$ and $a \leq b$. Such a set is *closed and bounded.*

Theorem 12.3.1 *If* $f : K \to \mathbb{R}$ *is a continuous function on a nonempty closed and bounded set* K *then* f *attains a maximum value and a minimum value on this set* K. *Thus, there exist points* $c, d \in K$ *such that*

$$\begin{aligned} f(c) &= \inf_{x \in K} f(x) \\ f(d) &= \sup_{x \in K} f(x). \end{aligned} \tag{12.6}$$

Chapter 13

Tangents, Slopes and Derivatives

The geometric notion of tangent is most easily understood for circles. A line is tangent to a circle at a point if that is the only point where the line and the circle meet. Another definition uses more geometry: a line is tangent to the circle through the point P, with center C, is it is perpendicular to the radius CP.

Both of these ideas are illuminating and reflect our intuition of what a tangent line ought to be. However, neither notion works very well for other curves. For example, visually it is perfectly clear that the line $y = 1$ is tangent to the graph $y = \sin x$, yet it meets this graph at infinitely many points. On the other hand any 'vertical' line meets $y = \sin x$ at just one point and yet such a line is surely not tangent to the graph. Since the graph $y = \sin x$ has no natural notion of 'center', it is also useless to try to define tangent as a line perpendicular to a 'radius.'

A geometrically elegant formulation of the notion of tangent arises naturally for the case of *ellipses.* Think of a circle C, and a tangent line l, at a point P, to the circle, drawn on a transparent sheet of paper. When a light is shown on the sheet from an angle, from a flashlight, the shadow C' cast on a wall by the circle C is a stretched out version of the circle. This curve C' is called an *ellipse.* The shadow l' cast by the tangent line l is again a straightline and is surely the tangent line to the ellipse C' at the point P', which is the shadow of P. This notion goes back to the greek study of *conic sections.*

Elegant though it is, even the method of the preceding paragraph fails to provide a definition of tangent that works for more general curves. For a general curve C we need to view a tangent line at a point P as a limiting form of the PQ (where Q is a 'nearby' point on C) as Q approaches P. This is formalized in the next section.

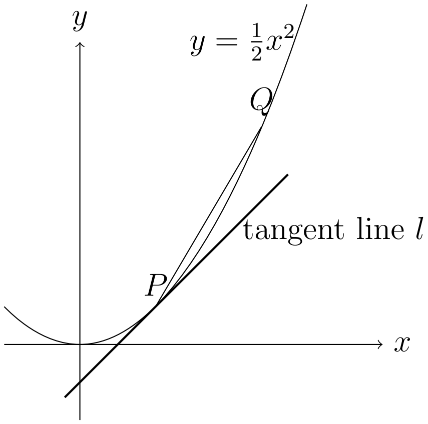


Figure 13.1: Tangent line and secant segment.

13.1 Secants and tangents

Consider a function f defined on $\mathbb{R}$ and think of the *graph* of f:

$$\{(x, y) \,:\, y = f(x), x \in \mathbb{R}\}.$$

Consider now a point

$$P = (x_*, y_*)$$

on this graph.

A *secant* line is a straight line through P and any other point Q on the graph.

Now think of all secant lines PQ, where $Q = (x, y)$ runs over points on the graph with x lying in some neighborhood U of x_*. There are possibly many such lines, each with a different slope: $(y - y_*)/(x - x*)$.

We shall say that a line l is *tangent* to the graph of f at the point P if it is the *unique* line that passes through P and has slope lying between the sups and infs of slopes of all 'nearby' secant lines:

$$\inf_{x \in U, x \neq x_*} \frac{f(x) - f(x_*)}{x - x_*} \leq \text{slope}(l) \leq \sup_{x \in U, x \neq x_*} \frac{f(x) - f(x_*)}{x - x_*} \tag{13.1}$$

for all neighborhoods U of x_*. If f is defined only on a subset S of $\mathbb{R}$ then we modify this definition by replacing $x \in U$ with $x \in U \cap S$. We interpet an infinite value, ∞ or $-\infty$, of slope to mean that the tangent line is 'vertical', parallel to the y-axis.

Another way to view the uniqueness of tangent line is to observe that this means that the slope of the tangent line is

$$\text{slope of tangent at point } P = \lim_{x \to x_*} \frac{f(x) - f(x_*)}{x - x_*}, \tag{13.2}$$

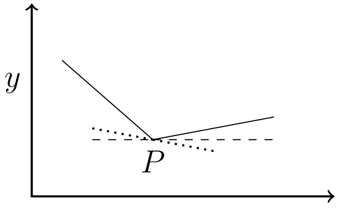
with the existence of this limit signifying the existence of a tangent line. *The tangent to the graph $y = f(x)$ at $P(x_*, f(x_*))$ is the line through P with slope given by (13.2).*

The slope of the tangent line to $y = f(x)$ at a point P is also called *the slope of the curve $y = f(x)$.*

For some functions there may be multiple lines l with slope satisfying the condition (13.1). For example, for the graph of

$$y = |x|$$

any line with slope $\in [-1, 1]$ satisfies the condition (13.1). Here is an illustration of a graph with a whole range of slopes satisfying the condition (13.1) at the point P:



Occasionally we will consider such a 'quasi-tangent' line to a graph. Though this is not a standard notion, let us agree that by a *quasi-tangent line* at $P(p, f(p))$ to the graph $y = f(x)$ for a function f we mean a line through P of slope satisfying the bounds

$$\inf_{w \in U \cap S, w \neq p} \frac{f(w) - f(p)}{w - p} \leq \text{slope of } l \leq \sup_{w \in U \cap S, w \neq p} \frac{f(w) - f(p)}{w - p}, \tag{13.3}$$

for every neighborhood U of p, where S is the domain of the function f. (Note that, as with tangent lines, this notion is meaningful only when $p \in S$ is not an isolated point of S.)

Thus, $y = f(x)$ has a tangent line at P if and only if it has a *unique* quasi-tangent line, and in this case the quasi-tangent line is the tangent line.

13.2 Derivative

Consider a function f defined on a set $S \subset \mathbb{R}$ and let p be a point of S that is not an isolated point. The *derivative* of f *at* p is defined to be:

$$f'(p) = \lim_{x \to p} \frac{f(x) - f(p)}{x - p}. \tag{13.4}$$

Thus *the derivative $f'(p)$ is the slope of the tangent to the graph $y = f(x)$ at the point $\bigl(p, f(p)\bigr)$*. Of course, if the graph fails to have a tangent line then it fails to have a derivative.

Let us look at some simple examples. First consider the constant function K whose value everywhere is 5:

$$K(x) = 5 \qquad \text{for all } x.$$

Common sense tells us that the slope of this is 0. We can check this readily from the official definition

$$\lim_{x \to p} \frac{K(x) - K(p)}{x - p} = \lim_{x \to p} \frac{0}{x - p} = \lim_{x \to p} 0 = 0.$$

We can elevate this observaton slightly by observing that we don't need K be equal to 5 *everywhere*, but just on a neighborhood of p.

If the function f is constant near p, then $f(x) = f(p)$, for x in a neighborhood of p, and so the derivative $f'(p)$ is 0. This just says that the graph is flat. Thus,

> *If a function is constant on a neighborhood of a point p then the derivative of the function at p is 0.*

Next, consider the function

$$g(x) = x \qquad \text{for all } x \in \mathbb{R}.$$

Then for any real number p we have

$$\lim_{x \to p} \frac{g(x) - g(p)}{x - p} = \lim_{x \to p} \frac{x - p}{x - p} = \lim_{x \to p} 1 = 1.$$

Hence the slope of

$$y = x$$

is 1: surely this is geometrically utterly obvious.

We can check readily that for the function $x \mapsto Mx + C$, where M and C are real numbers (constants) the derivative is M everywhere:

$$\lim_{x\to p} \frac{(Mx+C)-(Mp+C)}{x-p} = \lim_{x\to p} \frac{Mx-Mp}{x-p} = \lim_{x\to p} M = M.$$

13.3 Notation

The derivative if f at p is denoted

$$f'(p).$$

This is good for theoretical proofs and such but not very useful for practical algebraic calculations.

If a formula is given for $f(x)$ we denote the derivative of f at x by

$$\frac{df(x)}{dx} = f'(x).$$

The beginner's error in this notation is to put in a value for x in $df(x)/dx$: that is wrong usage of the notation:

$$\tfrac{df(3)}{d3} \text{ is wrong notation.}$$

Instead we should write

$$\tfrac{df(x)}{dx} \text{ at } x = 3, \text{ or } \tfrac{df(x)}{dx}\Big|_{x=3}$$

If we are writing $y = f(x)$ then the derivative of f at x is

$$\frac{dy}{dx}.$$

This notation meshes well with the derivative being the limit of the ratio of the increase in y to the increase in x:

$$\frac{dy}{dx} = \lim_{\Delta x\to 0} \frac{\Delta y}{\Delta x}, \tag{13.5}$$

where

$$\Delta y = y_Q - y_P, \quad \Delta x = x_Q - x_P,$$

with P being the point $(x,y) = (x, f(x))$ and $Q(w, f(w))$, and the limit $Q \to P$ is encoded in $\Delta x \to 0$.

For algebraic calculations the notation $f'(p)$ is inconvenient. Instead we use the notation

$$\frac{df(x)}{dx}$$

to denote the derivative of f at x. For example, for the function s given by

$$s(x) = x^2 \qquad \text{for all } x \in \mathbb{R}$$

we denote the derivative $s'(x)$ by

$$s'(x) = \frac{dx^2}{dx}.$$

Note that we don't mean that this is an actual ratio, nor do we mean that somehow the 'denominator' is d times x. The entire expression dx^2/dx shoud be viewed (at least at this stage) as one object, the derivative.

13.4 The derivative of x^2

Let us work out the derivative of the function given by $f(x) = x^2$ for all $x \in \mathbb{R}$ at $x = 3$. This is just the slope of the tangent line to the graph

$$y = x^2$$

at the point $(3, 3^2)$:

The slope of the secant line PQ is

$$\text{slope } PQ = \frac{x^2 - 3^2}{x - 3}$$

To find the slope of the tangent we need to let Q approach P; this means $x \to 3$, and we are looking then at the limit

$$\lim_{x \to 3} \frac{x^2 - 3^2}{x - 3}.$$

We can work this out easily. (Warning: avoid the 0/0 trap!) We factor $x^2 - 3^2$ as a product

$$x^2 - 3^2 = (x - 3)(x + 3)$$

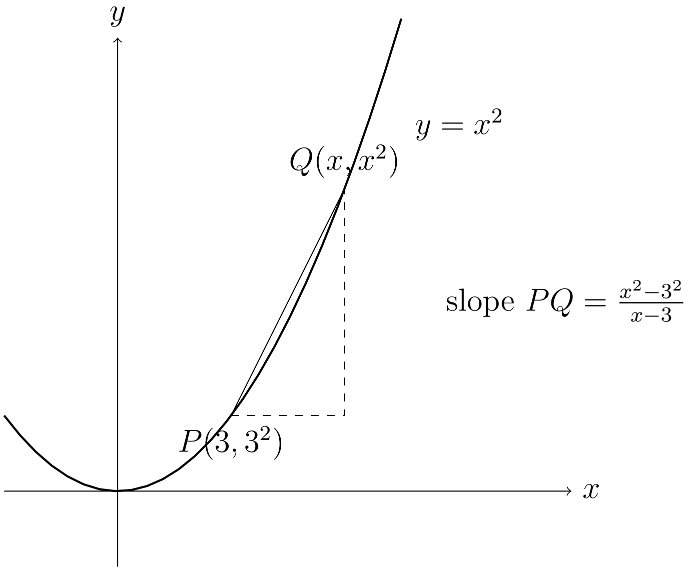


Figure 13.2: A secant segment to $y = x^2$ at $P(3, 3^2)$.

and obtain

$$\begin{aligned}\lim_{x\to 3}\frac{x^2-3^2}{x-3} &= \lim_{x\to 3}\frac{(x-3)(x+3)}{x-3}\\ &= \lim_{x\to 3}(x+3)\\ &= 6.\end{aligned} \tag{13.6}$$

Thus *the slope of the tangent to* $y = x^2$ *at the point* $(3, 3^2)$ *is* 6.

If you trace through the calculations above for a general point (p, p^2) on $y = x^2$ you see that the slope of the tangent at (p, p^2) is $2p$:

$$\begin{aligned}\lim_{x\to p}\frac{x^2-p^2}{x-p} &= \lim_{x\to p}\frac{(x-p)(x+p)}{x-p}\\ &= \lim_{x\to p}(x+p)\\ &= 2p.\end{aligned} \tag{13.7}$$

Thus the derivative of the function given by $f(x) = x^2$ at $x = p$ is $2p$. Using the notation $df(x)/dx$ this is displayed as

$$\frac{dx^2}{dx} = 2x. \tag{13.8}$$

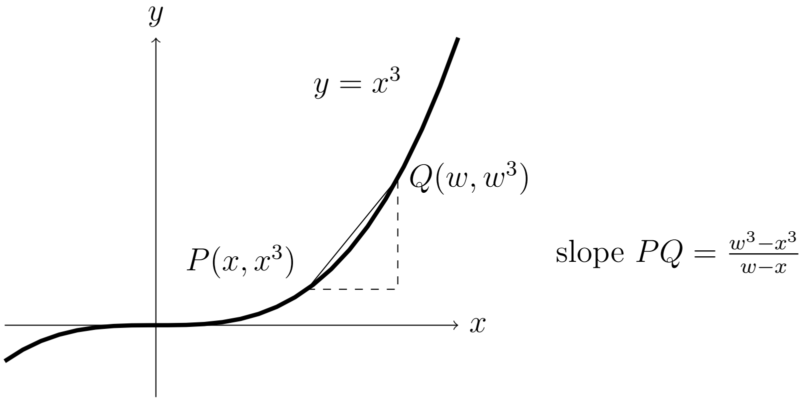


Figure 13.3: Secant segment for $y = x^3$ at $P(x, x^3)$.

13.5 Derivative of x^3

Let us do the calculation of the derivative for the function $f(x) = x^3$. Following the method used for $y = x^2$ we have first the picture

We can see that

$$\text{slope of } PQ = \frac{w^3 - x^3}{w - x}.$$

Letting $Q \to P$ makes the secant line PQ approach the tangent line at P in the limit. The slope of the tangent at P is then

$$\text{slope of tangent at } P = \lim_{w \to x} \frac{w^3 - x^3}{w - x}.$$

This just the derivative at x:

$$\begin{aligned}
\frac{dx^3}{dx} &= \lim_{w \to x} \frac{w^3 - x^3}{w - x} \\
&= \lim_{w \to x} \frac{(w - x)(w^2 + wx + x^2)}{w - x} \\
&\qquad \text{(using } A^3 - B^3 = (A - B)(A^2 + AB + B^2)\text{)} \qquad (13.9) \\
&= \lim_{w \to x} (w^2 + wx + x^2) \\
&= x^2 + x^2 + x^2 \\
&= 3x^2
\end{aligned}$$

13.6 Derivative of x^n

The procedure used for x^2 and x^n works for x^n, where n is any positive integer. This leads to

$$\frac{dx^n}{dx} = nx^{n-1}. \tag{13.10}$$

Thus, for example, the slope of the curve

$$y = x^7$$

at the point $(1, 1)$ is

$$7 \cdot 1^6 = 7.$$

13.7 Derivative of $x^{-1} = 1/x$

For $1/x$ we have the graph in Figure 13.4.

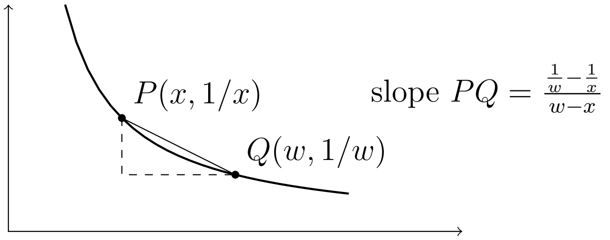


Figure 13.4: Secant segment for $y = 1/x$ at $P(x, 1/x)$.

The derivative at x is the slope of the tangent at $P(x, 1/x)$:

$$\frac{d\frac{1}{x}}{dx} = \lim_{Q \to P} (\text{slope of } PQ).$$

The slope of PQ is:

$$\text{slope of } PQ = \frac{(1/w) - (1/x)}{w - x}.$$

Then we can calculate the derivative as follows:

$$\begin{aligned}
\frac{d(1/x)}{dx} &= \lim_{w\to x} \frac{(1/w)-(1/x)}{w-x} \\
&= \lim_{w\to x} \frac{(x-w)/(xw)}{w-x} \\
&\qquad \left(\text{using } \tfrac{1}{A}-\tfrac{1}{B}=\tfrac{B-A}{AB}\right) \\
&= \lim_{w\to x} \frac{x-w}{xw(w-x)} \\
&= \lim_{w\to x} \frac{-1}{xw} \\
&= -\frac{1}{x^2}.
\end{aligned} \tag{13.11}$$

Thus:

$$\frac{d(1/x)}{dx} = -\frac{1}{x^2}.$$

Observe that this follows the formula $dx^n/dx = nx^{n-1}$:

$$\frac{dx^{-1}}{dx} = -1\cdot x^{-2},$$

even though $n=-1$ is not a positive integer.

The negative sign on $-1/x^2$ indicates a downward sloping tangent.

13.8 Derivative of $x^{-k} = 1/x^k$

Let k be a positive integer and consider the function

$$x^{-k} = \frac{1}{x^k}.$$

We can calculate its derivative:

$$\begin{aligned}\frac{d(1/x^k)}{dx} &= \lim_{w\to x}\frac{(1/w^k)-(1/x^k)}{w-x}\\ &= \lim_{w\to x}\frac{(x^k-w^k)/(x^kw^k)}{w-x}\\ &\qquad \text{(using } \tfrac{1}{A}-\tfrac{1}{B}=\tfrac{B-A}{AB}\text{)}\\ &= \lim_{w\to x}\frac{x^k-w^k}{x^kw^k(w-x)}\\ &= \lim_{w\to x}(-1)\cdot\frac{w^k-x^k}{w-x}\cdot\frac{1}{x^kw^k}\\ &= (-1)\cdot kx^{k-1}\cdot\frac{1}{x^{2k}},\end{aligned}\tag{13.12}$$

where in the last step we used the derivative of x^k:

$$\lim_{w\to x}\frac{w^k-x^k}{w-x}=kx^{k-1}.$$

Thus

$$\frac{d(1/x^k)}{dx}=-k\frac{1}{x^{2k-k+1}}=-k\frac{1}{x^{k+1}}.\tag{13.13}$$

Writing n for $-k$ this reads

$$\frac{dx^n}{dx}=nx^{n-1},$$

correct again, even though n is now a negative integer.

13.9 Derivative of $x^{1/2}=\sqrt{x}$

Consider the function

$$s(x)=\sqrt{x}=x^{1/2}$$

defined on all $x\geq 0$. Consider any $p\geq 0$. Then the derivative of this function at p is the slope of the tangent at $P(p,\sqrt{p})$ to the graph $y=\sqrt{x}$, and we know that

$$\text{slope of tangent at } P=\lim_{Q\to P}(\text{slope of } PQ).$$

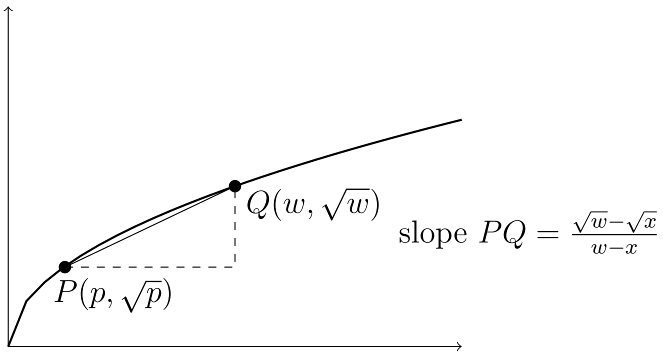


Figure 13.5: Secant segment for $y = \sqrt{x}$ at $P(p, \sqrt{p})$.

$$\begin{aligned} s'(p) &= \lim_{w \to p} \frac{\sqrt{w} - \sqrt{p}}{w - p} \\ &= \lim_{w \to p} \frac{(\sqrt{w} - \sqrt{p})(\sqrt{w} + \sqrt{p})}{(w - p)(\sqrt{w} + \sqrt{p})} \\ &\qquad \text{(using } (A - B)(A + B) = A^2 - B^2\text{)} \\ &= \lim_{w \to p} \frac{w - p}{wp(\sqrt{w} + \sqrt{p})} \\ &= \lim_{w \to p} \frac{1}{2(\sqrt{w} + \sqrt{p})} \\ &= \frac{1}{2\sqrt{p}}, \end{aligned} \tag{13.14}$$

when $p > 0$. If $p = 0$ we can just set $p = 0$ in the preceding calculations except for the very last line, being careful to note that the values w we work with are > 0 (because $s(w)$ is not defined for $w < 0$; thus,

$$s'(0) = \lim_{w \to 0+} \frac{1}{2(\sqrt{w} + \sqrt{0})} = \infty.$$

Thus the derivative of $\sqrt{x}$ is

$$\frac{d\sqrt{x}}{dx} = \frac{1}{2\sqrt{x}}, \tag{13.15}$$

when $x > 0$, and is ∞ when $x = 0$.

Note that

$$\sqrt{x} = x^{1/2} \quad \text{and} \quad \frac{1}{\sqrt{x}} = x^{-1/2},$$

and so the right side in (13.15) is

$$\frac{1}{2}x^{-1/2} = \frac{1}{2}x^{\frac{1}{2}-1}.$$

Thus

$$\frac{dx^{1/2}}{dx} = \frac{1}{2}x^{\frac{1}{2}-1}, \tag{13.16}$$

again agreeing with

$$\frac{dx^n}{dx} = nx^{n-1},$$

now with $n = 1/2$.

13.10 Derivatives of powers of x

Let $r = p/q$ be a rational number, where p and q are integers, and $q \neq 0$. Then for any $x \geq 0$ the power $x^{p/q}$ is the non-negative real number whose q-th power is x^p:

$$(x^{p/q})^q = x^p.$$

The existence of such a real number x^r follows from the intermediate value theorem, and ultimately is a consequence of the completeness of the real line.

The following derivative formula holds:

$$\frac{dx^r}{dx} = rx^{r-1}, \tag{13.17}$$

and can be proved by extension of the methods used before for negative powers of x and for $x^{1/2}$.

13.11 Derivatives with infinities

It does not seem useful to bother defining the derivative $F'(\infty)$, for a function F, defined at ∞. A natural extension of the geometric intuition of the derivative in terms of tangent lines leads to the definition

$$F'(p) = \lim_{x\to p} \frac{F(x)}{x} \qquad \text{if } p \in \{-\infty, \infty\}. \tag{13.18}$$

If $F(x)$ approaches a finite limit $F(\infty)$, as $x \to \infty$, then $F'(p)$ is 0, which conforms to intuition: the tangent line at $x = \infty$ is the 'horizontal' line $y = F(\infty)$.

Chapter 14

Derivatives of Trigonometric Functions

In this chapter we work out the derivative of sin, cos, and tan, by using their algebraic properties and the fundamental limits

$$\lim_{\theta\to 0}\frac{\sin\theta}{\theta}=1 \quad\text{and}\quad \lim_{\theta\to 0}\frac{\tan\theta}{\theta}=1.$$

14.1 Derivative of $\sin$ is $\cos$

The derivative of the funtion sin at $x \in \mathbb{R}$ is the slope of the graph of

$$y=\sin x$$

at the point $P(x,\sin x)$. Thus it is:

$$\sin' x=\lim_{w\to x}(\text{slope of } PQ),$$

where Q is the point $(w,\sin w)$.

Now the slope of PQ is

$$\text{slope of } PQ \; = \frac{\sin w-\sin x}{w-x}.$$

We have then

$$\sin' x=\lim_{w\to x}\frac{\sin w-\sin x}{w-x}.$$

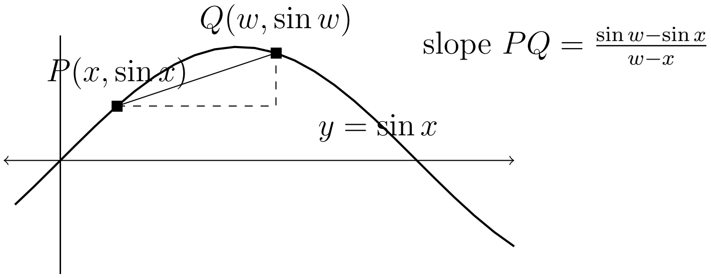


Figure 14.1: Secant segment for $y = \sin x$ at $P(x, \sin x)$.

To work this out there are two possible routes. We follow one, using the relation (8.22), which implies:

$$\sin w - \sin x = 2\sin\frac{w-x}{2}\cos\frac{w+x}{2}. \tag{14.1}$$

Using this we have

$$\begin{aligned} \sin' x &= \lim_{w\to x}\frac{2\sin\frac{w-x}{2}\cos\frac{w+x}{2}}{w-x} \\ &= \lim_{w\to x}\frac{2\sin\frac{w-x}{2}}{w-x}\cos\frac{w+x}{2}. \end{aligned} \tag{14.2}$$

To make this look more like something involving $\sin\theta/\theta$ (whose limit we understand), we write this as

$$\begin{aligned} \sin' x &= \lim_{w\to x}\frac{2\sin\frac{w-x}{2}}{2\frac{w-x}{2}}\cos\frac{w+x}{2} \\ &= \lim_{w\to x}\frac{\sin\frac{w-x}{2}}{\frac{w-x}{2}}\cos\frac{w+x}{2} \\ &= 1\cdot\cos\frac{x+x}{2}. \end{aligned} \tag{14.3}$$

Thus we have found the derivative of sin:

$$\frac{d\sin x}{dx} = \sin' x = \cos x. \tag{14.4}$$

14.2 Derivative of $\cos$ is $-\sin$

The derivative of the funtion cos at $x \in \mathbb{R}$ is

$$\cos' x = \lim_{w \to x} \frac{\cos w - \cos x}{w - x}.$$

Recall the relation (8.23):

$$\cos w - \cos x = -2 \sin \frac{w - x}{2} \sin \frac{w + x}{2}. \tag{14.5}$$

Using this we have

$$\begin{aligned} \cos' x &= -\lim_{w \to x} \frac{2 \sin \frac{w-x}{2} \sin \frac{w+x}{2}}{w - x} \\ &= -\lim_{w \to x} \frac{2 \sin \frac{w-x}{2}}{w - x} \cos \frac{w + x}{2} \\ &= -\lim_{w \to x} \frac{2 \sin \frac{w-x}{2}}{2 \frac{w-x}{2}} \sin \frac{w + x}{2} \\ &= -\lim_{w \to x} \frac{\sin \frac{w-x}{2}}{\frac{w-x}{2}} \sin \frac{w + x}{2} \\ &= -1 \cdot \sin \frac{x + x}{2}. \end{aligned} \tag{14.6}$$

Thus we have found the derivative of sin:

$$\frac{d \cos x}{dx} = \cos' x = -\sin x. \tag{14.7}$$

14.3 Derivative of $\tan$ is $\sec^2$

The derivative of the funtion tan at $x \in \mathbb{R}$ is

$$\tan' x = \lim_{w \to x} \frac{\tan w - \tan x}{w - x}.$$

Recall the relation (8.24):

$$\tan(a - b) = \frac{\tan a - \tan b}{1 + \tan a \tan b}.$$

From this we have

$$\tan a - \tan b = (1 + \tan a \tan b)\tan(a - b).$$

Taking w for a and x for b we have

$$\tan w - \tan x = (1 + \tan w \tan x)\tan(w - x).$$

We can proceed to the derivative now:

$$\begin{aligned}
\tan' x &= \lim_{w \to x} \frac{\tan w - \tan x}{w - x} \\
&= \lim_{w \to x} \frac{(1 + \tan w \tan x)\tan(w - x)}{w - x} \\
&= \lim_{w \to x} (1 + \tan w \tan x)\frac{\tan(w - x)}{w - x} \\
&= (1 + \tan x \tan x) \cdot 1,
\end{aligned} \tag{14.8}$$

because

$$\lim_{\theta \to 0} \frac{\tan \theta}{\theta} = 1.$$

Thus,

$$\tan' x = 1 + \tan^2 x.$$

Now recall that $1 + \tan^2 x$ is $\sec^2 x$. Hence

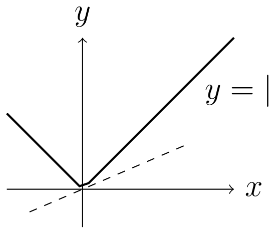
$$\frac{d \tan x}{dx} = \tan' x = \sec^2 x. \tag{14.9}$$

Chapter 15

Differentiability and Continuity

15.1 Differentiability implies continuity

We know that not all continuous functions are differentiable. For example, the absoute value function is continuous but is not differentiable at 0:



However, if a function is differentiable then it is continuous:

Theorem 15.1.1 *Suppose* f *is a function defined on a set* $S \subset \mathbb{R}$ *and is differentiable at a point* $p \in S$. *Then* f *is ccontinuous at* p.

Proof. Recall that the derivative $f'(p)$ is given by

$$f'(p) = \lim_{x \to p} \frac{f(x) - f(p)}{x - p}.$$

Note that this is meaningful only when $p \in S$ is not an isolated point of S. We wish to show that $f(x) \to f(p)$ when $x \to p$. For this we first write $f(x)$ as $f(p)$ plus the amount it deviates from $f(p)$:

$$f(x) = f(p) + f(x) - f(p).$$

Since we have information about $\frac{f(x)-f(p)}{x-p}$ let us bring this in:

$$f(x) = f(p) + (x-p)\left[\frac{f(x)-f(p)}{x-p}\right].$$

Now let $x \to p$:

$$\lim_{x\to p} f(x) = f(p) + 0 \cdot f'(p) = f(p),$$

where we used the given fact that $f'(p)$ exists and is *finite* (the application of the 'limit of product equals product of limits' argument would fail if $f'(p)$ were ∞ or $-\infty$). $\boxed{\text{QED}}$

We have allowed infinite values for derivatives. For example, the function f specified by

$$f(x) = \begin{cases} -1 & \text{if } x < 0; \\ 0 & \text{if } x = 0; \\ 1 & \text{if } x > 0, \end{cases}$$

has derivative

$$f'(0) = \infty.$$

But note that f is *not* continuous at 0.

There are plenty of functions that are *continuous everywhere and yet differentiable nowhere.* They are hard to visualize and have severely zig-zag graphs.

Chapter 16

Using the Algebra of Derivatives

In this chapter we use the basic algebraic rules for working with derivatives. In summary the basic algebraic rules are

$$
\begin{aligned}
\frac{d(U+V)}{dx} &= \frac{dU}{dx} + \frac{dV}{dx} \\[2ex]
\frac{d(kU)}{dx} &= k\frac{dU}{dx} \\
\frac{d(UV)}{dx} &= \frac{dU}{dx}V + U\frac{dV}{dx} \qquad \text{(product rule)} \\[2ex]
\frac{d\frac{1}{V}}{dx} &= -\frac{1}{V^2}\frac{dV}{dx} \\[2ex]
\frac{d\frac{U}{V}}{dx} &= \frac{V\frac{dU}{dx} - U\frac{dV}{dx}}{V^2} \qquad \text{(quotient rule)}
\end{aligned}
\tag{16.1}
$$

where U and V are differentiable functions for which all the quantities on the right sides exist and are finite, and k is any constant (real number). We have retained some redundancy here; for example, that $d(kU)/dx = kdU/dx$ can be deduced from the 'product rule' for $d(UV)/dx$, keeping in mind that $dk/dx = 0$.

16.1 Using the sum rule

The sum rule is so easy to use that it is best forgotten that we are actually using a rule (that needs to be proved). For example

$$\begin{aligned}\frac{d(5\sin x+4x^3-3)}{dx} &= \frac{d(5\sin x)}{dx}+\frac{d(4x^3-3)}{dx}\\ &\quad \text{(assuming these derivatives exist and are finite)}\\ &= 5\frac{d\sin x}{dx}+\frac{d(4x^3)}{dx}+\frac{d(-3)}{dx}\\ &\quad \text{(assuming these derivatives exist and are finite)}\\ \\ &= 5\cos x+4\frac{dx^3}{dx}+0\\ &= 5\cos x+12x^2.\end{aligned}$$

For this very first example we were careful to state most of the steps and logic, but there is no need to be so extreme. Mostly we can write down the derivative of a sum directly:

$$\frac{d(5\sin x-3x^5+2x-4)}{dx}=5\cos x-15x^4+2.$$

16.2 Using the product rule

The product rule gets us to more complicated functions. Take for example,

$$y=\sqrt{x}\sin x.$$

Using the product rule we find the derivative very easily:

$$\frac{dy}{dx}=\frac{d\sqrt{x}}{dx}\sin x+\sqrt{x}\frac{d\sin x}{dx}=\frac{1}{2\sqrt{x}}\sin x+\sqrt{x}\cos x.$$

We can use the product rule for multiple products:

$$\begin{aligned}\frac{d(UVW)}{dx} &= \frac{dUV(W)}{dx}\\ &= \frac{dU}{dx}VW+U\frac{dVW}{dx} \qquad (16.2)\\ &= \frac{dU}{dx}VW+U\frac{dV}{dx}W+UV\frac{dW}{dx}.\end{aligned}$$

Clearly this patterns works for any number of products.

Here is another example of the product rule

$$\begin{aligned}\frac{d(t+\tan t)(2t^2+\sqrt{t})(4t-\cos t)}{dt}\\
&= \frac{d(t+\tan t)}{dt}(2t^2+\sqrt{t})(4t-\cos t)+(t+\tan t)\frac{(2t^2+\sqrt{t})}{dt}\\
&\qquad + (t+\tan t)(2t^2+\sqrt{t})\frac{d(4t-\cos t)}{dt}\\
&= (1+\sec^2 t)(2t^2+\sqrt{t})(4t-\cos t)+(t+\tan t)(4t+\frac{1}{2\sqrt{t}})(4t-\cos t)\\
&\qquad + (t+\tan t)(2t^2+\sqrt{t})(4+\sin t).\end{aligned}$$

16.3 Using the quotient rule

Let us first see quickly how to use the formula

$$(1/V)' = -\frac{1}{V^2}V'.$$

Here is an example

$$\begin{aligned}\frac{d}{dw}\left(\frac{1}{\sin w+\cos w+1}\right) &= -\frac{1}{(\sin w+\cos w+1)^2}\frac{d(\sin w+\cos w+1)}{dw}\\
&= -\frac{1}{(\sin w+\cos w+1)^2}(\cos w-\sin w).\end{aligned}$$

Now let us do an example of the full quotient rule

$$(U/V)' = \frac{VU'-UV'}{V^2}.$$

We have

$$\begin{aligned}\frac{d}{dw}\left(\frac{w+\tan w}{\sin w+\cos w+1}\right) &= \frac{(\sin w+\cos w+1)\frac{d(w+\tan w)}{dw}-(w+\tan w)\frac{d(\sin w+\cos w+1)}{dw}}{(\sin w+\cos w+1)^2}\\
&= \frac{(\sin w+\cos w+1)(1+\sec^2 w)-(w+\tan w)(\cos w-\sin w)}{(\sin w+\cos w+1)^2}.\end{aligned}$$

Chapter 17

Using the Chain Rule

The chain rule, along with the algebraic rules we have already studied, makes it possible to work out derivatives of functions given by highly complicated expressions.

17.1 Initiating examples

Before stating the chain rule in the abstract we can see its essence in a few basic examples. As a first example, look at

$$\frac{d\sin(x^2+3x)}{dx} = \left[\cos(x^2+3x)\right](2x+3)$$

The sin has become cos (which is the derivative of sin), and then we have a multiplier $2x+3$ which we recognize to be the derivative of x^2+3x.

Next look at

$$\frac{d(4t^3-2\sin t)^{1/3}}{dt} = \frac{1}{3}(4t^3-2\sin t)^{1/3-1}\cdot(12t^2-2\cos t).$$

We recognize the factor on the right as the derivative of the function $(\cdot)^{1/3}$, evaluated at $4t^3-2\sin t$; next, this is multiplied by the derivative of $4t^3-2\sin t$.

As our last initiating example, look at

$$\frac{d\sin\left(\sqrt{x^4+x^2}\right)}{dx} = \left[\cos\left(\sqrt{x^4+x^2}\right)\right]\cdot\frac{1}{2\sqrt{x^4+x^2}}\cdot(4x^3+2x).$$

This is a chain rule applied twice: first the sin is differentiated to obtain $\cos\bigl(\cdot\bigr)$, next $\sqrt{\cdot}$ is differentiated to produce $\frac{1}{2\sqrt{\cdot}}$, and, finally, x^4+x^2 is differentiated to produce the last factor $4x^3+2x$.

17.2 The chain rule

Consider a function H that is the *composite* of functions F and G:

$$H(x) = F\bigl(G(x)\bigr).$$

This means that to calculate the value $H(x)$ we must first work out the value

$$G(x)$$

and then apply the function F to it:

$$F\bigl(G(x)\bigr).$$

For example, the function given by

$$\sqrt{x^4+x^2}$$

is the composite of the square root function $\sqrt{\cdot}$ with the function given by x^4+x^2.

As another example,

$$\sin\sqrt{w}$$

is the cmomposite of sin with $\sqrt{\cdot}$.

The *composite*

$$H = F\circ G$$

is the function whose value is given by

$$H(t) = F\bigl(G(t)\bigr),$$

for every value t for which $G(t)$ is defined and $F\bigl(G(t)\bigr)$ is also defined. For example, the composite function given by

$$\sqrt{1+x}$$

is defined only for $x \geq -1$.

The chain rule says that if

$$H = F \circ G$$

then

$$H'(x) = F'\big(G(x)\big)G'(x) \tag{17.1}$$

provided that values and derivatives on the right exist and are finite.

Returning to examples, we have then

$$\frac{d\tan\sqrt{x}}{dx} = \left[\sec^2\sqrt{x^4+3x^2}\right]\frac{1}{2\sqrt{x}},$$

because we recognize that

$$\tan\sqrt{x}$$

is obtained by applying tan to $\sqrt{x}$:

$$\tan\sqrt{x} = (\tan\circ\sqrt{\ })(x).$$

Chapter 18

Proving the Algebra of Derivatives

We now explore precise statements and proofs for the rules of algebra for derivatives.

18.1 Sums

If f and g are functions with domain $S \subset \mathbb{R}$ then their sum is the function $f + g$ on S whose value at any $x \in S$ is given by

$$(f + g)(x) = f(x) + g(x).$$

Suppose now that $p \in S$ is a point where the derivatives $f'(p)$ and $g'(p)$ exist. Then

$$(f + g)'(p) = f'(p) + g'(p) \tag{18.1}$$

if this sum is defined (that is not $\infty + (-\infty)$ or $(-\infty) + \infty$).

This result follows directly from the fact that the limit of a sum is the sum of the limits.

18.2 Products

If f and g are functions with domain $S \subset \mathbb{R}$ then their product is the function fg on S whose value at any $x \in S$ is given by

$$(fg)(x) = f(x)g(x).$$

For proving the product rule we have to bring in a useful little geometric observation about products. Notice that if a rectangle, whose sides are A and B, gets enlarged so that it becomes a C by D rectangle then its area increases by

$$\begin{aligned} CD - AB &= CD - AD + AD - AB \\ &= (C-A)D + A(D-B) \\ &= (C-A)(D-B+B) + A(D-B) \\ &= (C-A)(D-B) + (C-A)B + A(D-B). \end{aligned} \tag{18.2}$$

Proposition 18.2.1 *Let f and g be functions on a set $S \subset \mathbb{R}$, and at $p \in S$ is a point where f and g are both differentiable (that is, the derivatives $f'(p)$ and $g'(p)$ exist and are finite). Then*

$$(fg)'(x) = f'(x)g(x) + f(x)g'(x). \tag{18.3}$$

Proof. Recall that for any function h the derivative $h'(p)$ is defined to be

$$h'(x) = \lim_{w \to x} \frac{h(w) - h(x)}{w - x}.$$

Let us apply this to $h = fg$. Then

$$(fg)'(w) = \lim_{w \to x} \frac{f(w)g(w) - f(x)g(x)}{w - x}. \tag{18.4}$$

Now we split the numerator following the idea of (18.2):

$$\begin{aligned} f(w)g(w) - f(x)g(x) = &\,[f(w) - f(x)]\,[g(w) - g(x)] \\ &+ [f(w) - f(x)]\,g(x) + f(x)\,[g(w) - g(x)]\,. \end{aligned}$$

Now divide by $w - x$ to obtain

$$\begin{aligned} \frac{f(w)g(w) - f(x)g(x)}{w - x} = &\,[f(w) - f(x)]\left[\frac{g(w) - g(x)}{w - x}\right] \\ &+ \left[\frac{f(w) - f(x)}{w - x}\right] g(x) + f(x)\left[\frac{g(w) - g(x)}{w - x}\right]. \end{aligned}$$

The ratio $\frac{f(w)-f(x)}{w-x}$ approaches the derivative $f'(x)$ when $w \to x$, and similarly for $\frac{g(w)-g(x)}{w-x}$, so we rewrite everything in terms of these 'difference quotients':

$$\begin{aligned}\frac{f(w)g(w)-f(x)g(x)}{w-x} &= (w-x)\left[\frac{f(w)-f(x)}{w-x}\right]\left[\frac{g(w)-g(x)}{w-x}\right] \\ &\quad + \frac{f(w)-f(x)}{w-x}g(x) + f(x)\frac{g(w)-g(x)}{w-x}.\end{aligned}$$

Now just let $w \to x$:

$$\begin{aligned}\lim_{w\to x}\frac{f(w)g(w)-f(x)g(x)}{w-x} &= 0\cdot f'(x)\cdot g'(x) \\ &\quad + f'(x)g(x) + f(x)g'(x),\end{aligned}$$

which works because the derivatives $f'(x)$ and $g'(x)$ have been assumed to exist and be finite. Thus fg is differentiable at x and

$$(fg)'(x) = f'(x)g(x) + f(x)g'(x). \qquad \boxed{\text{QED}}$$

18.3 Quotients

We turn now to proving the quotient rule:

$$(f/g)'(x) = \frac{g(x)f'(x) - f(x)g'(x)}{g(x)^2}, \tag{18.5}$$

valid whenever f and g are functions defined on some common domain S and x is a point of S where f and g are both differentiable and $g(x) \neq 0$.

From the definition of the derivative, $(f/g)'(x)$ is the limit of

$$\frac{\frac{f(w)}{g(w)} - \frac{f(x)}{g(x)}}{w-x} \tag{18.6}$$

as $w \to x$. Let us rework the numerator so it involves mainly the differences

$f(w) - g(w)$ and $g(w) - g(x)$:

$$\begin{aligned}
\frac{f(w)}{g(w)} - \frac{f(x)}{g(x)} &= \frac{f(w)g(x) - f(x)g(w)}{g(w)g(x)} \\
&= \frac{[f(w) - f(x) + f(x)]\, g(x) - f(x)\,[g(w) - g(x) + g(x)]}{g(w)g(x)} \\
&= \frac{[f(w) - f(x)]\, g(x) + f(x)g(x) - f(x)\,[g(w) - g(x)] - f(x)g(x)}{g(w)g(x)} \\
&= \frac{[f(w) - f(x)]\, g(x) - f(x)\,[g(w) - g(x)]}{g(w)g(x)}.
\end{aligned}$$

The preceding algebra may look very complicated at first but it has a natual and simple thinking behind it: at each stage where we see $f(w)$ we write it in terms of the difference $f(w) - f(x)$ as $f(w) - f(x) + f(x)$, and the same for $g(w)$.

Thus

$$\frac{\left(\frac{f(w)}{g(w)} - \frac{f(x)}{g(x)}\right)}{w - x} = \frac{\left[\frac{f(w)-f(x)}{w-x}\right] g(x) - f(x)\left[\frac{g(w)-g(x)}{w-x}\right]}{g(w)g(x)}. \tag{18.7}$$

Now we let $w \to x$. To deal with the denominator term $g(w)$, we need to use the fact the the differentiability of g at x makes it continuous as x:

$$g(w) \to g(x),$$

as $w \to x$. Applying this to the identity (18.7) produces

$$\lim_{w\to x} \frac{\left(\frac{f(w)}{g(w)} - \frac{f(x)}{g(x)}\right)}{w - x} = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)g(x)},$$

which is what we wished to prove. $\boxed{\text{QED}}$

If in the quotient rule we take the numerator to be the constant function 1 we obtain

$$(1/V)' = \frac{V \cdot 1' - 1 \cdot V'}{V^2} = \frac{0 - V'}{V^2} = -\frac{1}{V^2}V',$$

a useful formula in itself.

Chapter 19

Proving the Chain Rule

It is easy to understand why the chain rule works. But, as we shall see, turning this easy understanding into a proof runs into a snag. For the official proof we then follow a different line of reasoning.

19.1 Why it works

Consider the composite function $F \circ G$:

$$y = F\bigl(G(x)\bigr).$$

To work out the derivative dy/dx, let us introduce some notation:

$$y = F(u) \qquad \text{where} \qquad u = G(x).$$

Then

$$y = F(G(x)) = H(x),$$

and

$$\frac{\Delta y}{\Delta x} = \frac{\Delta y}{\Delta u}\,\frac{\Delta u}{\Delta x}. \tag{19.1}$$

Letting $\Delta x \to 0$ gives

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}. \tag{19.2}$$

For example, for the function

$$y = \sqrt{1 + \sin x},$$

we take $u = 1 + \sin x$, and then

$$y = \sqrt{u} \qquad \text{and} \qquad u = 1 + \sin x,$$

so that

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = \frac{1}{2\sqrt{u}}\cos x = \frac{1}{2\sqrt{1+\sin x}}\cos x.$$

This is clearly the chain rule as we have seen and used before.

The argument used above is natural but has one technical gap. What if the denominator $\Delta u = \Delta G(x)$ keeps hitting 0? Then the first ratio on the right in (19.1) is undefined and the argument breaks down.

There are two ways to deal with this road block. Either we can try to tread through the obstacle very carefully, or we can try an entirely different pathway, one that is less natural but one that gets us to the destination faster. We will go through such a proof in the next section. This is a situation one sometimes faces in trying to construct a proof out of a reasonable idea. In the end, however, the 'reasonable idea' often gives better insight into 'why' the result is true. (In fact, the idea does provide a proof in the case $G'(x)$ is not 0.)

19.2 Proof the chain rule

A proof of the chain rule can be built out of the following useful observation:

Lemma 19.2.1 *Let f be a function, with domain $S \subset \mathbb{R}$, differentiable at a point $p \in \mathbb{R}$. Then there is a function f_p on S, which is continuous at p, and for which*

$$f(x) = f(p) + (x-p)f_p(x) \qquad \text{for all } x \in S, \tag{19.3}$$

and

$$f_p(p) = f'(p).$$

There is a more enlightening way to state (19.3):

$$f(x) = f(p) + (x-p)f'(p) + \epsilon_p(x) \qquad \text{for all } x \in S, \tag{19.4}$$

where

$$\lim_{x \to p} \frac{\epsilon_p(x)}{x-p} = 0, \tag{19.5}$$

because $\epsilon_p(x) = (x-p)[f_p(x) - f'(p)]$. To understand the significance of (19.4) observe that the first two terms of the right describe the y-value of the tangent line to $y = f(x)$ at p, and so condition (19.4) says:

$$\lim_{x \to p} \frac{f(x) - T_p f(x)}{x-p} = 0, \tag{19.6}$$

where

$$y = T_p f(x) \overset{\text{def}}{=} f(p) + (x-p)f'(p)$$

is the equation of the tangent to the graph of f at $\big(p, f(p)\big)$.

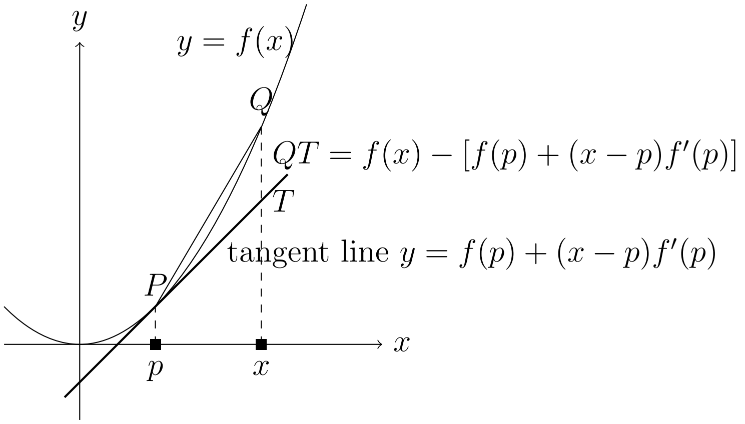

y
$y = f(x)$
Q
$QT = f(x) - [f(p) + (x-p)f'(p)]$
T
tangent line $y = f(p) + (x-p)f'(p)$
P
x
p
x

Figure 19.1: The tangent as an approximation to the graph

Proof. Simply note that

$$f(x) = f(p) \,+\, f(x) - f(p) = f(p) \,+\, (x-p)\left[\frac{f(x)-f(p)}{x-p}\right], \tag{19.7}$$

for all $x \in S$ with $x \neq p$. We want to denote the ratio $\frac{f(x)-f(p)}{x-p}$ by $f_p(x)$. However, we need to say what $f_p(p)$ is, for at $x = p$ the ration $\frac{f(x)-f(p)}{x-p}$ is not defined. But, we do know that as $x \to p$, the ration $\frac{f(x)-f(p)}{x-p}$ approaches $f'(p)$. So let us define the function f_p on S by

$$f_p(x) = \begin{cases} \frac{f(x)-f(p)}{x-p} & \text{if } x \in S \text{ and } x \neq p; \\ f'(p) & \text{if } x = p. \end{cases} \tag{19.8}$$

Then

$$\lim_{x\to p} f_p(x) = \lim_{x\to p} \frac{f(x) - f(p)}{x - p} = f'(p) = f_p(p),$$

which shows that f_p is continuous at p. Moreover, from (19.7),

$$f(x) = f(p) + (x - p)f_p(x)$$

for $x \in S$ wthe $x \neq p$. But putting in $x = p$ shows that this is also valid when $x = p$. $\boxed{\text{QED}}$

Now we can prove the chain rule.

Proposition 19.2.1 *Let f and g be functions on subsets of $\mathbb{R}$, and let S_0 be the set of all x for which the composite*

$$f \circ g : x \mapsto f\big(g(x)\big)$$

is defined. Suppose that p is a point of S_0 such that g is differentiable at p and f is differentiable at $g(p)$. Assume also that p is not an isolated point of S_0. Then the function $f \circ g$ is differentiable at p and

$$(f \circ g)'(p) = f'\big(g(p)\big)g'(p). \tag{19.9}$$

We can avoid all the worrying about domains if we simply assume $f \circ g$ is defined in a neighborhood of p. Thus, if $f \circ g$ is defined in a neighborhood of p, g is differentiable at p and f is differentiable at $g(p)$ then $f \circ g$ is differentiable at p and (19.9) holds.

Proof. Let

$$q = g(p).$$

Recall from Lemma 19.2.1 the functions f_q and g_p. Then

$$\begin{aligned} f\big(g(x)\big) &= f(q) + \big(g(x) - q\big)f_q\big(g(x)\big) \\ &= f(q) + \big(q + (x - p)g_p(x) \, - q\big)f_q\big(g(x)\big) \\ &= f(q) + (x - p)g_p(x)f_q\big(g(x)\big) \end{aligned} \tag{19.10}$$

for all x in the domain of $f \circ g$. Then

$$\frac{f\big(g(x)\big) - f\big(g(p)\big)}{x - p} = g_p(x)f_q\big(g(x)\big)$$

for all $x \in S_0$ with $x \neq p$. Now letting $x \to p$ (recall that p has been assumed to be a limit point of S_0) we have

$$\begin{aligned}\lim_{x\to p} \frac{f\big(g(x)\big) - f\big(g(p)\big)}{x-p} &= \lim_{x\to p} g_p(x) f_q\big(g(x)\big) \\ &= \underbrace{g_p(p)}_{g'(p)}\, \underbrace{f_q(g(p))}_{f'\big(g(p)\big)}\end{aligned} \tag{19.11}$$

where in the last step we used several observations and facts: (i) because g_p is continuous at p the limit $\lim_{x\to p} g_p(x)$ is $g_p(p)$, (ii) f_q is continuous at q and g is continuous at p (because g is differentiable at p) and so $\lim_{x\to p} f_q\big(g(x)\big) = f_q(g(p))$; (iii) both $g_p(p) = g'(p)$ and $f_q(g(p)) = f_q(q) = f'(q)$ are finite (real numbers).

Chapter 20

Using Derivatives for Extrema

Derivative can be used to find maximum and minimum values of functions. For example, we will see soon how to find maximum and minimum values of

$$g(x) = 2x^3 - 6x^2 + 1$$

of $x \in [-1, 1]$.

For a function f on a closed interval $[a, b]$, where $a, b \in \mathbb{R}$ and $a < b$, here is the strategy for finding maximum and minimum values:

(i) work out the derivative $f'(x)$ and find the values of x in the interval $[a, b]$ where $f'(x)$ is 0;

(ii) the maximum (minimum) of f is the largest (smallest) of the values of f at the points where f' is 0 and the endpoint values $f(a)$ and $f(b)$,

assuming that f is continuous on $[a, b]$ and is differentiable in the interior (a, b) (or at least at those points in the interior where f reaches maximum/minimum value).

To see how this works let us apply it to the function $g(x) = 2x^3 - 6x^2 + 1$ for $x \in [-1, 1]$. The derivative is

$$g'(x) = 6x^2 - 12x = 6x(x - 2).$$

This is 0 at $x = 0$ and at $x = 2$. Snce $x = 2$ is outside the domain $[-1, 1]$, we ignore it. Now we compute the values of g at $x = 0$ and at the endpoints 1 and -1:

$$g(0) = 1 \qquad g(1) = -3, \qquad g(-1) = -7.$$

The largest value of $g(x)$ is therefore 1, occurring at $x = 0$, and the smallest value is -7, occurring at $x = -1$.

The proof that this strategy works is postponed to Chapter 21.

If the function f is defined on the whole real line $\mathbb{R}$ or intervals such as $(2, \infty)$, we have to modify step (ii) above to:

(ii)' the supremum (infimum) of a function f, defined on an interval U, is the largest (smallest) of the values of f at the points where f' is 0 and the endpoint limit values $\lim_{x\to a} f(a)$ and $\lim_{x\to b} f(b)$, where a and b are the endpoints of the interval U and the limits here are assumed to exist.

If the limits $\lim_{x\to a} f(x)$ and $\lim_{x\to b} f(b)$ don't exist then, of course, this method doesn't work.

The term 'maximum' is used when the supremum is actually attained at a point in the domain of the function; similarly, we speak of the 'minimum' value of a function if the infimum is actually attained in the domain of the function. For example, $1/x$, for $x \in (0, \infty)$, has no maxima or minima but its supremum is ∞ (as $x \downarrow 0$) and its infimum is 0 (as $x \to \infty$).

20.1 Quadratics with calculus

Let us apply the method of calculus to find the minimum value of the quadratic function

$$y(x) = 3x^2 - 6x + 16$$

The derivative is

$$6x - 6$$

and this is 0 when $x = 1$. The value of y here is

$$y(1) = 13.$$

There are no boundary points given, and our function is defined on the entire real line $\mathbb{R}$. So we need to work out the endpoint limits:

$$\lim_{x\to\infty} y(x) = \infty \qquad \text{and} \qquad \lim_{x\to-\infty} y(x) = \infty.$$

Thus

$$\sup_{x\in\mathbb{R}} y(x) = \infty,$$

and

$$\inf_{x\in\mathbb{R}} y(x) = y(1) = 13.$$

20.2 Quadratics by algebra

There is a way to obtain the minimum value of a quadratic function by pure algebra, with no use of calculus. This is by using the ancient method of 'completing the square':

$$3x^2-6x+16 = 3(x^2-2x)+16 = 3(x^2-2x+1-1)+16 = 3(x^2-2x+1)-3+16$$

which shows that

$$3x^2-6x+16 = 3(x-1)^2+13.$$

The first term on the right is always ≥ 0, with minimum value 0 when $x = 1$. Hence $3x^2-6x+16$ has minimum value

$$0+13 = 13,$$

and this value is attained when $x = 1$.

Let us now look at the general quadratic

$$Ax^2+Bx+C,$$

with A, B, C being real numbers, with $A \neq 0$. Completing the square we have

$$\begin{aligned}
Ax^2+Bx+C &= A\left[x^2+\frac{B}{A}x\right]+C \\
&= A\left[x^2+2\frac{B}{2A}x+\left(\frac{B}{2A}\right)^2-\left(\frac{B}{2A}\right)^2\right]+C \\
&= A\left[x^2+2\frac{B}{2A}x+\left(\frac{B}{2A}\right)^2\right]-A\left(\frac{B}{2A}\right)^2+C \qquad (20.1) \\
&= A\left[x+\frac{B}{2A}\right]^2-\frac{B^2}{4A}+\frac{4AC}{4A} \\
&= A\left[x+\frac{B}{2A}\right]^2-\frac{(B^2-4AC)}{4A}
\end{aligned}$$

Thus,

$$Ax^2 + Bx + C = A\left[x + \frac{B}{2A}\right]^2 - \frac{(B^2 - 4AC)}{4A} \tag{20.2}$$

If $A > 0$ then the first term on the right is always ≥ 0, with minimum value occurring at $x = -B/(2A)$:

$$\min_{x\in\mathbb{R}}(Ax^2 + Bx + C) = -\frac{(B^2 - 4AC)}{4A} \quad \text{if } A > 0. \tag{20.3}$$

If A is negative then the first term on the right in (20.2) is always ≤ 0 and the largest it gets is 0, this happening when $x = -B/(2A)$; so

$$\max_{x\in\mathbb{R}}(Ax^2 + Bx + C) = -\frac{(B^2 - 4AC)}{4A} \quad \text{if } A < 0. \tag{20.4}$$

This is a clean and nice solution, but in practice it is faster to simply observe that

$$(Ax^2 + Bx + C)' = 2Ax + B$$

is 0 when $x = -B/(2A)$ and this point corresponds to the maximum/minimum value of $Ax^2 + Bx + C$.

The classic use of the method of completing the square is in obtaining the solutions of the quadratic equation

$$Ax^2 + Bx + C = 0 \tag{20.5}$$

Using the completed square form this reads

$$A\left[x + \frac{B}{2A}\right]^2 - \frac{(B^2 - 4AC)}{4A} = 0$$

from which we have

$$\left[x + \frac{B}{2A}\right]^2 = \frac{(B^2 - 4AC)}{4A^2}.$$

Taking square roots shows that

$$x + \frac{B}{2A} = \pm\frac{\sqrt{B^2 - 4AC}}{2A},$$

where ± signifies that there are two choices. Thus the two solutions of the quadratic equation (20.5) are

$$\alpha = \frac{-B + \sqrt{B^2 - 4AC}}{2A} \qquad \text{and} \qquad \beta = \frac{-B - \sqrt{B^2 - 4AC}}{2A}.$$

Observe that

$$\alpha - \beta = \frac{\sqrt{B^2 - 4AC}}{A}.$$

The square of this is

$$(\alpha - \beta)^2 = \frac{B^2 - 4AC}{A^2}.$$

The quantity

$$A^2(\alpha - \beta)^2 = B^2 - 4AC \tag{20.6}$$

is called the *discriminant* of the quadratic

$$Ax^2 + Bx + C.$$

If the discriminant is 0 then (20.6) shows that $\alpha = \beta$. On the other hand if the discriminant is not 0 then the roots α and β are distinct.

If the discriminant is < 0 then looking at the expressions for α and β we see that they are not real numbers (since square roots of negatives are involved).

20.3 Distance to a line

We work out the distance of a point

$$P(x_P, y_P)$$

from a line L:

$$y = mx + k.$$

This distance is, by definition, the shortest distance from P to any point on the line:

$$d(P, L) \overset{\text{def}}{=} \inf_{Q \in L} d(P, Q),$$

where

$$d(P, Q) = \text{distance between the points } P \text{ and } Q.$$

If Q has coordinates (x, y) then

$$d(P, Q) = \sqrt{(x - x_P)^2 + (y - y_P)^2}.$$

We have to find the minimum value of this as (x, y) runs over the line L. We can avoid unpleasant calculations by minimizing the distance squared:

$$d(P, Q)^2 = (x - x_P)^2 + (y - y_P)^2. \tag{20.7}$$

Clearly if we can find the minimum value of this then we can just take the square-root to find the minimum distance. Keep in mind that in (20.7) (x, y) is on the line L and so

$$y = mx + k.$$

If we write out $d(P, Q)^2$ in terms of x we have

$$d(P, Q)^2 = (x - x_P)^2 + (mx + k - y_P)^2.$$

This is clearly quadratic in x:

$$\begin{aligned} d(P, Q)^2 &= x^2 - 2x_P x + x_P^2 + m^2 x^2 + 2m(k - y_P)x + (k - y_P)^2 \\ &= (1 + m^2)x^2 + 2[-x_P + (k - y_P)]x + x_P^2 + (k - y_P)^2. \end{aligned}$$

The coefficient of x^2 is

$$1 + m^2,$$

which is positive. From our study of quadratic functions we know then that $d(P, Q)^2$ does attain a minimum value and at the point Q_0 where it attains minimum the derivative

$$\frac{d}{dx} d(P, Q)^2$$

is 0.

The derivative of $d(P, Q)^2$ with respect to x is:

$$\frac{d}{dx} d(P, Q)^2 = 2(x - x_P) + 2(y - y_P)\frac{dy}{dx} = 2[(x - x_P) + (y - y_P)m].$$

This is 0 if and only if (x, y) is the special point

$$Q_0(x_0, y_0)$$

which satisfies

$$(x_0 - x_P) + (y_0 - y_P)m = 0. \tag{20.8}$$

It is worth observing that this implies

$$\frac{y_0 - y_P}{x_0 - x_P} = -\frac{1}{m}, \tag{20.9}$$

assuming that $m \neq 0$ and that P isn't actually on the line L.

The geometric significance of (20.9) is that the line PQ_0 has slope $-1/m$, and this means that

$$PQ_0 \text{ is perpendicular to the line } L.$$

Geometrically this makes perfect sense.

Now returning to (20.8), we subsitute in the value of y as $mx+k$ to obtain

$$(x_0 - x_P) + (mx_0 + k - y_P)m = 0,$$

which is

$$(1 + m^2)x_0 + km - x_P - y_P m = 0.$$

Solving this we obtain the following value for x:

$$x_0 = \frac{x_P + y_P m - km}{1 + m^2}. \tag{20.10}$$

The corresponding value for y is

$$\begin{aligned} y_0 &= mx_0 + k \\ &= m\left(\frac{x_P + y_P m - km}{1 + m^2}\right) + k \\ &= \frac{mx_P + y_P m^2 - km^2}{1 + m^2} + \frac{k(1 + m^2)}{1 + m^2} \\ &= \frac{mx_P + m^2 y_P + k}{1 + m^2} \end{aligned}$$

Thus

$$y_0 = \frac{mx_P + m^2 y_P + k}{1 + m^2} \tag{20.11}$$

We have thus found the point

$$Q_0(x_0, y_0)$$

on the line L that is closest to the point P.

We can now work out the distance between P and Q_0:

$$\begin{aligned} d(P,Q_0)^2 &= (x_0-x_P)^2+(y_0-y_P)^2 \\ &= [-m(y_0-y_P)]^2+(y_0-y_P)^2 \quad \text{on using (20.8)} \\ &= m^2(y_0-y_P)^2+(y_0-y_P)^2 \\ &= (m^2+1)(y_0-y_P)^2 \end{aligned} \tag{20.12}$$

We need to work oit $y_0 - y_P$ from (20.11):

$$\begin{aligned} y_0-y_P &= \frac{mx_P+m^2y_P+k}{1+m^2}-\frac{(1+m^2)y_P}{1+m^2} \\ &= \frac{mx_P+m^2y_P+k-y_P-m^2y_P}{1+m^2} \\ &= \frac{mx_P+k-y_P}{1+m^2} \end{aligned}$$

Using this in the formula (20.12) for $d(P,Q_0)^2$ we have:

$$d(P,Q_0)^2=(m^2+1)\left(\frac{mx_P+k-y_P}{1+m^2}\right)^2=\frac{(mx_P+k-y_P)^2}{1+m^2}$$

Taking the square root produces at last the distance of P from the line L:

$$d(P,L)=\frac{|mx_P+k-y_P|}{\sqrt{1+m^2}}. \tag{20.13}$$

Taking a concrete example, let us work out

the distance of the point $(1,2)$ from the line $y = 5x - 2$.

This works out to

$$\frac{|5*1-2\ -\ 2|}{\sqrt{1+5^2}}=\frac{1}{\sqrt{26}}.$$

The absolute value in the numerator erases a piece of information. In this example,

$$5*1-2>2$$

and this means that the point $(1,2)$ lies *below* the line $y = 5x - 2$.

In fact,

$$mx_P+k\ -y_P$$

measures how far 'below' (in the vertical y-direction) the line L the point P lies. If α is the angle between L and the positive x-axis then

$$m = \tan\alpha$$

and so

$$\frac{1}{\sqrt{1+m^2}} = \frac{1}{\sqrt{1+\tan^2\alpha}} = \frac{1}{\sqrt{\sec^2\alpha}} = |\cos\alpha|.$$

Thus

$$\frac{|mx_P + k - y_P|}{\sqrt{1+m^2}} = |(mx_P + k - y_P)\cos\alpha|.$$

If you view this geometrically it is clear that this does indeed measure the distance between P and the line L.

Now consider a different way of writing the equation of a line:

$$Ax + By + C = 0,$$

where at least one of A and B is not 0. Assume that $B \neq 0$ (if B were 0 then the line would be 'vertical', parallel to the y-axis). Then we can rewrite the equation as

$$y = -\frac{A}{B}x + \frac{-C}{B}.$$

So, to switch back to our previous notation,

$$m = -\frac{A}{B}, \quad k = \frac{-C}{B}.$$

Then the distance between $P(x_P, y_P)$ and L is

$$\frac{|mx_P + k - y_P|}{\sqrt{1+m^2}} = \frac{|-\frac{A}{B}x_P - \frac{C}{B} - y_P|}{\sqrt{1+\frac{A^2}{B^2}}}$$

Simplifying the algebra this produces the formula

$$d(P,L) = \frac{|Ax_P + By_P + C|}{\sqrt{A^2+B^2}}. \tag{20.14}$$

You can check that this formula works even when B is 0, for then the line L has constant x value $-C/A$ and the x-coordinate of P is x_P, so that the distance is

$$|x_P - (-C/A)| = \frac{|Ax_P + C|}{|A|},$$

which matches (20.14) for $B = 0$.

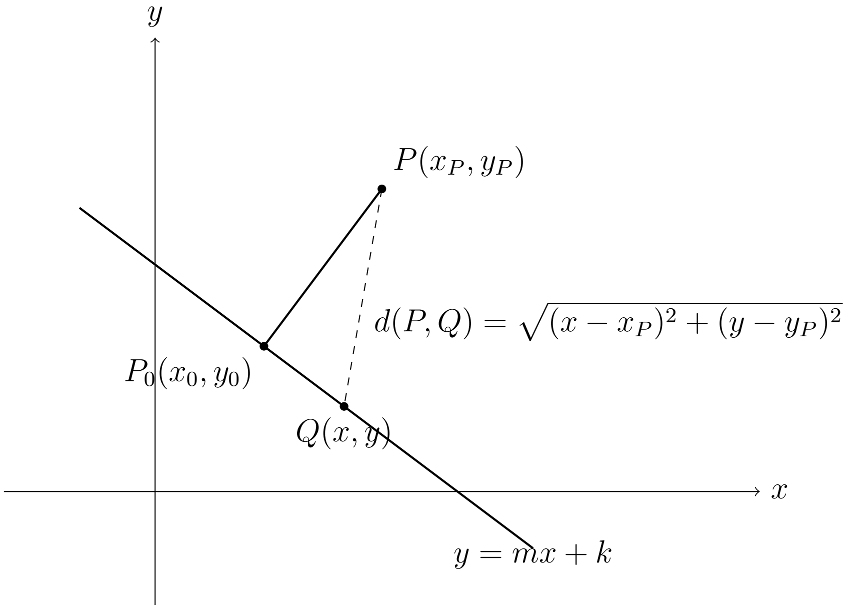


Figure 20.1: Distance of a point P from points on a line.

20.4 Other geometric examples

A straight piece of wire of length L units is to be cut into two pieces, one of which will be bent into a square and the other a circle. What is the maximum and what is the minimum possible total area (enclosed by the square and circle) that can be enclosed in this way?

Intuition suggest that the largest area would be obtained if we take the entire wire and bend it into a circle. This intuition (where does it come from?) is verified to be correct by the mathematical solution we work out. It is not clear intuitively how to cut the wire to obtain the minimum wire.

Let x units be the length of the piece that is bent into a circle. Thus if the radius of the circle is R then

$$2\pi R = x,$$

and so the area enclosed by the circle is

$$\pi R^2 = \pi \left(\frac{x}{2\pi}\right)^2 = \frac{x^2}{4\pi}.$$

The remaining piece is of length

$$L - x$$

and when this is bent to form a square, each side of the square has length

$$\frac{L-x}{4},$$

and its area is

$$\left(\frac{L-x}{4}\right)^2.$$

Thus the total area enclosed is

$$A(x) = \frac{1}{4\pi}x^2 + \frac{1}{16}(L-x)^2. \tag{20.15}$$

We have to find the maximum and minimum values of $A(x)$, keeping in mind that x cannot be negative or more than L:

$$x \in [0, L].$$

(Taking $x = 0$ means we just form a large square and no circle, and taking $x = L$ means we form a circle out of the full length of wire and no square at all.)

The derivative of A is:

$$A'(x) = \frac{1}{4\pi}2x + \frac{1}{16}2(L-x)\cdot(-1) = \frac{x}{2\pi} + \frac{x-L}{8} = \frac{4x + \pi x - \pi L}{8\pi},$$

which simplifies to

$$A'(x) = \frac{(4+\pi)x - \pi L}{8\pi}.$$

Thus

$$\text{the solution of } A'(x) = 0 \text{ is } x_0 = \tfrac{\pi}{\pi+4}L.$$

The remaining length $L - x$ to be bent into a square is:

$$L - x_0 = L - \frac{\pi}{\pi+4}L = \frac{4}{\pi+4}L.$$

The total area enclosed is

$$A(x_0) = \frac{1}{4\pi}\left(\frac{\pi}{\pi+4}L\right)^2 + \frac{1}{16}\left(\frac{4}{\pi+4}L\right)^2 \tag{20.16}$$

Simplifying this we have

$$\begin{aligned} A(x_0) &= \frac{\pi^2}{4\pi(\pi+4)^2}L^2 + \frac{16}{16(\pi+4)^2}L^2 \\ &= \frac{\pi}{4(\pi+4)^2}L^2 + \frac{1}{(\pi+4)^2}L^2 \\ &= \frac{\pi}{4(\pi+4)}L^2 + \frac{4}{4(\pi+4)^2}L^2 \\ &= \frac{\pi+4}{4(\pi+4)^2}L^2 \\ &= \frac{1}{4(\pi+4)}L^2 \end{aligned} \tag{20.17}$$

We compare this value

$$A(x_0) = \frac{L^2}{4\pi+16}$$

with the endpoint values obtained from (20.15) with $x = 0$ and $x = L$:

$$A(0) = \frac{L^2}{16} \qquad \text{and} \qquad A(L) = \frac{L^2}{4\pi}.$$

Among these values $A(L)$ is the highest, having the smallest denominator 4π, and $A(x_0)$ is the least, having the largest denominator $4\pi + 16$.

Thus the largest area is enclosed when $x = L$, which means we take the entire length of wire and bend it into a circle.

Now consider another problem. A rectangle of sides L units by W units has four little squares, each having side x units, cut out of the four corners; the edges are now folded to form a box (with no cover). The height of the box is x units, and the edges are $L - 2x$ units and $W - 2x$ units. What should x be to maximize the volume of the box?

The volume V cubic units is given by

$$V(x) = x(L-2x)(W-2x) = xLW - 2(L+W)x^2 + 4x^3$$

The value of x is ≥ 0 but cannot be more than $W/2$ (we assume W is the shorter edge, that is: $W \leq L$). Thus,

$$x \in [0, W/2].$$

The derivative $V'(x)$ is

$$V'(x) = LW - 4(L+W)x + 12x^2$$

For this to be 0 we have

$$12x^2 - 4(L+W)x + LW = 0.$$

The solutions are

$$\frac{4(L+W) \pm \sqrt{16(L+W)^2 - 4 * 12 * LW}}{24} = \frac{4(L+W) \pm \sqrt{16[(L+W)^2 - 3LW]}}{24}$$

which simplify to

$$\frac{(L+W) \pm \sqrt{L^2 + W^2 - LW}}{6}$$

We need to check if these two values fall within the interval $[0, W/2]$. Since $L \geq W$ the numerator for the $+$ sign is

$$\begin{aligned}(L+W) + \sqrt{L^2 + W^2 - LW} &= L + W + \sqrt{L(L-W) + W^2} \\ &\geq W + W + \sqrt{0 + W^2} \\ &= 3W\end{aligned}$$

which makes the ratio

$$\frac{(L+W) + \sqrt{L^2 + W^2 - LW}}{6} \geq \frac{3W}{6} = \frac{W}{2},$$

falling outside the allowed range, unless we have the extreme case $L = W$ for which the ratio is $W/2$. On the other hand, if we take the $-$ sign, then the numerator is

$$(L+W) - \sqrt{(L-W)^2 + LW}$$

and the term being subtracted is larger than $\sqrt{(L-W)^2} = L - W$, and so

$$(L+W) - \sqrt{(L-W)^2 + LW} \leq (L+W) - (L-W) = 2W,$$

and so then the ratio is

$$\frac{(L+W) - \sqrt{L^2 + W^2 - LW}}{6} \leq \frac{2W}{6} = \frac{W}{3}.$$

Thus

$$x_0 = \frac{(L+W) - \sqrt{L^2 + W^2 - LW}}{6}$$

is in the interior of $[0, W/2]$. Clearly this choice of x must produce the maximum value of the volume, for the value of $V(x)$ at the endpoints $x = 0$ and $x = W/2$ is 0.

Thus the maximum volume is

$$V(x_0) = x_0(L - 2x_0)(W - 2x_0)$$

After a long calculation this works out to

$$\frac{1}{54}\left[(L+W)(5LW - 2L^2 - 2W^2) + 2(L^2 + W^2 - LW)\sqrt{L^2 + W^2 - LW}\right].$$

If we start with a square, for which $L = W$, this simplifies to

$$\frac{2}{27}L^3,$$

with x_0 being $L/6$.

Exercises on Maxima and Minima

1. Find the maximum and minimum values of x^2 for $x \in [-1, 2]$.

2. Find the maximum and minimum values of

   $$x(6-x)(3-x)$$

   for $x \in [0, 2]$.

3. A wire of length 12 units is bent to form an isosceles triangle. What should the lengths of the sides of the triangle be to make its area maximum?

4. A piece of wire is bent into a rectangle of maximum area. Show that this maximal area rectangle is a square.

5. A piece of wire of length L is cut into pieces of length x and $L - x$ (including the possibilty that x is 0 or L), and each piece is bent into a circle. What is the value of x which would make the total area enclosed by the pieces maximum, and what is the value of x which would make this area minimum.

6. Here are some practice problems on straight lines and distances:

   (i) Work out the distance from $(1, 2)$ to the line $3x = 4y + 5$

   (ii) Work out the distance from $(2, -2)$ to the line $4x - 3y - 5 = 0$.

   (iii) Find the point P_0 on the line L, with equation $3x + 4y - 7 = 0$, closest to the point $(0, 3)$. What is the angle between P_0P and the line L?

   (iv) Let P_0 be the point on the line L, with equation $3x + 4y - 11 = 0$, closest to the point $P(1, 3)$. What is the slope of the line P_oP?

   (v) Let P_0 be the point on the line L, with equation $3x + 4y - 11 = 0$, closest to the point $P(1, 3)$. Find the equation of the line through P and P_0.

7. Prove the inequality

$$\frac{x^3}{3} + \frac{k^{3/2}}{3/2} \geq kx, \tag{20.18}$$

   for all $x, k \in (0, \infty)$. Explain when $\geq$ is $=$. [Hint: Show that, for any fixed value $k \in (0, \infty)$, the maximum value of

$$\Phi(x) = kx - \frac{x^3}{3} \qquad \text{for } x \in (0, \infty)$$

   is $\frac{k^{3/2}}{3/2}$. Note that $\Phi(0) = 0$ and $\lim_{x\to\infty} \Phi(x) = -\infty$; so you have to find a point $p \in (0, \infty)$ where $\Phi'(p)$ is 0 and compare the value $\Phi(p)$ with $\Phi(0)$ and choose the larger.]

8. Prove the inequality

$$x^6 + 5k^{6/5} \geq 6kx, \tag{20.19}$$

   for all $x, k \in (0, \infty)$. Now show that

$$x^6 + 5y^6 \geq 6y^5x,$$

   for all $x, y \in (0, \infty)$.

Chapter 21

Local Extrema and Derivatives

21.1 Local Maxima and Minima

Consider a function f on a set $S \subset \mathbb{R}$. Suppose $u \in S$ is such that the value of f at u is maximum:

$$f(u) = \sup_{x \in S} f(x).$$

Assume that this point u acually lies in the *interior* of S:

$$u \in S^0.$$

Then, in any neighborhood U of u, there are points of S to the right of u that are in U and there are points of S to the left of d that are also in U.

Then for any $x \in U$ to the *right* of u we have

$$\frac{f(x) - f(u)}{x - u} \geq 0$$

because both numerator and denominator of the fraction on the left are ≥ 0. On the other hand, for $x \in U$ to the *left* of u we have

$$\frac{f(x) - f(u)}{x - u} \leq 0$$

because the numerator is ≥ 0 but the denominator is < 0. Thus, on any neighborhood U of u, the sup of the ratio $\frac{f(x)-f(u)}{x-u}$ is ≥ 0 and its inf is ≤ 0:

$$\inf_{x \in U, x \neq u} \frac{f(x) - f(u)}{x - u} \leq 0 \leq \sup_{x \in U, x \neq u} \frac{f(x) - f(u)}{x - u}.$$

But this just means that the line of slope 0, through the point $P = (u, f(u))$ on the graph of f, satisfies the condition (13.1) for quasi-tangents.

A very similar argument works for a point b where f is a minimum.

In fact these arguments easily establish:

Proposition 21.1.1 *Suppose f is a function on a set $S \subset \mathbb{R}$, and $b \in S$ is a point in the interior of S. Let U be a neighborhood of b contained in S and suppose*

$$f(b) \leq f(x) \qquad \text{for all } x \in U.$$

If, moreover, $f'(b)$ exists then

$$f'(b) = 0.$$

Suppose $u \in S$ is such that there is a neighborhood U of u with $U \subset S$ and

$$f(u) \geq f(x) \qquad \text{for all } x \in U.$$

If, moreover, $f'(u)$ exists then

$$f'(u) = 0.$$

For a function f defined on a set $S \subset \mathbb{R}$, a point $b \in S$ is said to be a *local minimum* if there is a neighborhood U of b on which the value of f at b is $\leq$ all other values:

$$f(b) \leq f(x) \quad \text{for all } x \in U \cap S.$$

A point u is said to be a *local maximum* of f if the value $f(u)$ is $\geq$ all other values over U:

$$f(u) \geq f(x) \quad \text{for all } x \in U \cap S.$$

With this terminology, Proposition 21.1.1 says that if a function has a local minimum or a local maximum in the *interior* of its domain of definition, and if the graph has a tangent line at that point, then the this tangent line is horizontal, that is, it has 0 slope.

Recalling the term 'quasi-tangent' we introduced back at the end of section (13.1) we see a sharper form of the preceding proposition:

> If f is a function, defined on a set S, that has a local minimum or a local maximum at a point p in the interior of S, then the line through p of zero slope is a quasi-tangent to the graph $y = f(x)$ at the point $\big(p, f(p)\big)$.

Review Exercises

1. For the set
$$S = [-\infty, -1) \cup (1, 2] \cup \{6, 8\} \cup [9, \infty]$$
write down
   (i) an interior point
   (ii) a limit point
   (iii) a boundary point
   (iv) an isolated point
   (v) the interior $S^0 =$
   (vi) the boundary $\partial S =$

2. Answer and explain briefly:
   (i) If $4 < \sup T$ is 4 an upper bound of T?
   (ii) In (i), is there a point of T that is > 4?
   (iii) If $\inf T < 3$ is 3 a lower bound of T?
   (iv) In (iii), is there a point of T that is < 3?

3. Answer the following concerning limits, with brief explanations:
   (i) If $\lim_{x\to 1} F(x) = 2$ does it follow that $F(1) = 2$?
   (ii) If g is continuous at 3 is g differentiable at 3?
   (iii) If g is differentiable at 5 is g continuous at 5?
   (iv) If $h'(5) = 4$ and $h(5) = 8$ then $\lim_{x\to 5} h(x) =$
   (v) If $H'(2) = 5$ and $H(2) = 3$ then $\lim_{w\to 2} \frac{H(w)-3}{w-2} =$
   (vi) If $G'(5) = 1$ and $G(5) = 6$ then $\lim_{y\to 5} \frac{G(y)-6}{y-5} =$
   (vii) $\lim_{w\to 0} \frac{\sin w}{w} =$
   (viii) $\lim_{w\to \pi/3} \frac{\sin w - \sin(\pi/3)}{w-\pi/3} =$

(ix) If $G'(3) = 4$ then

$$\lim_{h\to 0} \frac{G(3+h) - G(3)}{h} =$$

4. Work out the following derivatives:

   (i) $\frac{d\sqrt{w^4-2w^2+4}}{dw}$

   (ii) $\frac{d[(1+\sqrt{y})\tan y]}{dy}$

   (iii) $\frac{d\left[\frac{1+\sqrt{y})}{\tan y}\right]}{dy}$

   (iv) $\frac{d\cot x}{dx}$

   (v) $\frac{d\sin\left(\cos\left(\tan\left(\sqrt{x}\right)\right)\right)}{dx}$

5. Using the definition of the derivative, show that

$$\frac{d(1\sqrt{x})}{dx} = -\frac{1}{2x\sqrt{x}}.$$

Chapter 22

Mean Value Theorem

In this chapter we explore a very powerful result in calculus: the mean value theorem. This result shows that for any differentiable function f, the slope of a secant line PQ is actually equal to the slope of a suitable tangent line to the graph $y = f(x)$.

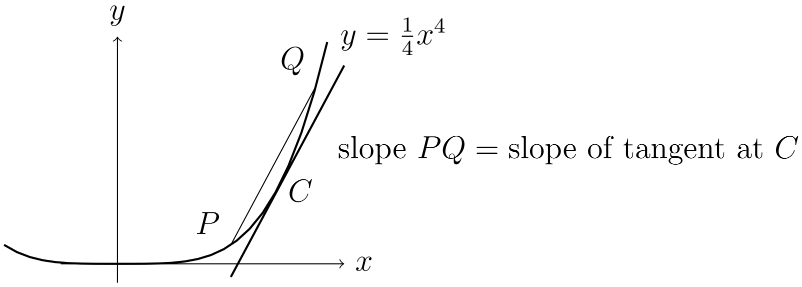


Figure 22.1: A tangent line parallel to a secant.

22.1 Rolle's Theorem

The following is a version of *Rolle's theorem*, which is a key step towards proving the mean value theorem:

Theorem 22.1.1 *Let f be a function, continuous on the interval $[a, b]$, where $a, b \in \mathbb{R}$ with $a < b$, and suppose*

$$f(a) = f(b).$$

Suppose also that the derivative $f'(x)$ exists for all $x \in (a,b)$. Then there is a point p on the graph of f over (a,b) where there is a tangent line of zero slope:

$$f'(p) = 0 \qquad \textit{for some } p \in (a,b).$$

Proof. Since f is continuous we know that there is a point in $[a,b]$ where it reaches its maximum value and a point in $[a,b]$ where it reaches its minimum value.

Let us first see that at least one of these values must occur at a point p in (a,b), the *interior* of the interval. If both the maximum and the minimum of f were to occur at the end points a and b, then, since $f(a) = f(b)$, the function f must be *constant*, say with value K; picking any $p \in (a,b)$ we have $f(p) = K$, which is both the maximum and the minimum of f. Thus in all cases there is a $p \in (a,b)$ such that f attains either its maximum or its minimum value at p. Then, by Proposition 21.1.1, $f'(p) = 0$. $\boxed{\text{QED}}$

There is a small sharpening of Rolle's theorem we could note, just to see how proofs can be tweaked to sharpen results. Recall that if f is continuous on $[a,b]$ and attains either a maximum or a minimum value $f(p)$ at $p \in [a,b]$, then there is a quasi-tangent line at $(p, f(p))$ to the graph $y = f(x)$ that is flat. Thus, we could drop the requirement in Rolle's theorem that f is differentiable and conclude that there is a point $p \in [a,b]$ where the graph$y = f(x)$ has a flat quasi-tangent line.

22.2 Mean Value Theorem

The following is the enormously useful *Mean Value Theorem*:

Theorem 22.2.1 *Let f be a function, continuous on the interval $[a,b]$, where $a, b \in \mathbb{R}$ with $a < b$. Suppose also that the derivative $f'(x)$ exists for all $x \in (a,b)$. Then there is a point on the graph of f over (a,b) where there is a tangent line that has slope equal to*

$$\frac{f(b) - f(a)}{b - a}.$$

Thus,

$$f'(p) = \frac{f(b) - f(a)}{b - a} \qquad \textit{for some } p \in (a,b). \tag{22.1}$$

Proof. Consider the secant line passing through the points $A = (a, f(a))$ and $B = (b, f(b))$. The slope of this line is

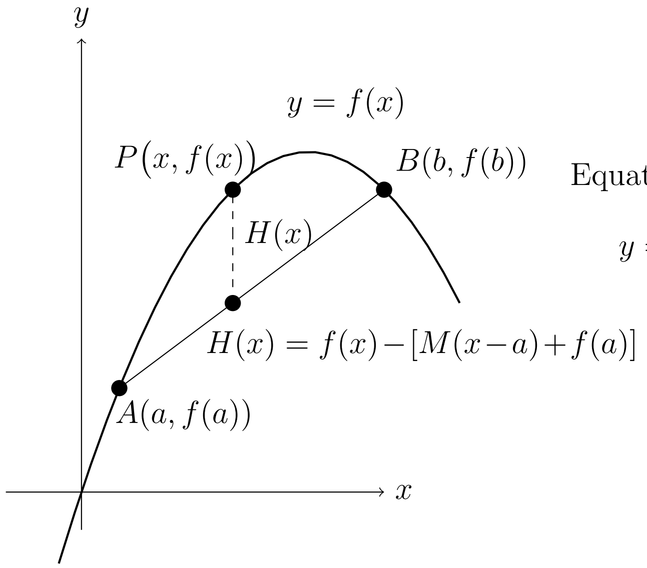
$$M = \frac{f(b) - f(a)}{b - a}. \tag{22.2}$$

Its equation is

$$y - f(a) = M(x - a).$$

Consider now how high f rises above this line:

$$H(x) = f(x) - [M(x - a) + f(a)] \qquad \text{for all } x \in [a, b]. \tag{22.3}$$



Equation of line AB is

$$y = M(x - a) + f(a)$$

Figure 22.2: The height H of the graph of f above a secant AB.

This function is continuous, being the sum of two continuous functions. Moreover,

$$H(a) = H(b) = 0.$$

Hence there is a point $p \in (a, b)$ where the tangent line to the graph of H is flat:

$$H'(p) = 0.$$

From the expression for $H(x)$ given in (22.3) we have

$$H'(x) = f'(x) - M.$$

(This makes geometric sense: the slope of H is the slope of f minus the slope of the line $M(x-a)+f(a)$.) So the relation $H'(p)=0$ means

$$f'(p)-M=0,$$

which means $f'(p)=M$. $\boxed{\text{QED}}$

22.3 Rolle's theorem on $\mathbb{R}^*$

The argument used to prove Rolle's theorem (Theorem 22.1.1) extends directly to cover the case of functions defined on subsets of $\mathbb{R}^*$:

Theorem 22.3.1 *Let $F:[a,b]\to\mathbb{R}^*$ be a continuous function, where $a,b\in\mathbb{R}^*$ with $a<b$, with $F(x)\in\mathbb{R}$ for all $x\in(a,b)$, and suppose*

$$F(a)=F(b).$$

Suppose also that the derivative $F'(x)$ exists for all $x\in(a,b)$. Then

$$F'(p)=0 \qquad \text{for some } p\in(a,b).$$

We will use this result in establishing one case of l'Hospital's rule (Proposition 28.2.2).

Chapter 23

The Sign of the Derivative

In this chapter we harness the power of the mean value theorem and a very precise understanding of the notion of limit (as explored in Proposition 10.1.1) to study the relationship between the sign (positive/negative/zero) of the derivative f' of a function and the nature of the function f, whether it is increasing, decreasing or constant.

Recall that for a function f defined on a set $S \subset \mathbb{R}$, and a point $p \in S$, if U is a neighborhood of p then part of U might not be inside S. So to work with values $f(x)$ for x in the neighborhood U we must focus on $x \in U \cap S$, which would guarantee that x does lie in the domain S of f.

23.1 Positive derivative and increasing nature

Intuitively it is clear that a function is increasing whereever its slope is ≥ 0, and it is decreasing wherever its slope is ≤ 0. In this section we make this idea precise.

The simplest observation on slopes and derivatives is that if f is an increasing function on an interval then its slope is ≥ 0:

Proposition 23.1.1 *Let f be a function on a set $S \subset \mathbb{R}$.*

If f is increasing on S, that is if $f(s) \leq f(t)$ for all $s, t \in S$ with $s \leq t$, then $f'(p) \geq 0$ for all $p \in S$ where $f'(p)$ exists.

If f is decreasing on S, that is if $f(s) \geq f(t)$ for all $s, t \in S$ with $s \leq t$, then $f'(p) \leq 0$ for all $p \in S$ where $f'(p)$ exists.

Proof. This follows directly from the definition of the derivative:

$$f'(p) = \lim_{x\to p} \frac{f(x) - f(p)}{x - p}.$$

If f is increasing then $f(x) \geq f(p)$ when $x > p$ (thus $x - p > 0$) in S and $f(x) \leq f(p)$ when $x < p$ (thus $x - p < 0$) in S. Hence the ratio $\frac{f(x)-f(p)}{x-p}$ is ≥ 0, and so the limit $f'(p)$ is also ≥ 0.

If f is decreasing then

$$\frac{f(x) - f(p)}{x - p} \leq 0$$

both when $x > p$ and when $x < p$, with $x \in S$. Hence in this case $f'(p) \leq 0$. $\boxed{\text{QED}}$

The following is a much sharper result going in the other direction:

Proposition 23.1.2 *Let f be a function on a set $S \subset \mathbb{R}$, and p a point in S where $f'(p)$ exists and is positive, that is*

$$f'(p) > 0.$$

Then there is a neighborhood U of p such that the $f(x) > f(p)$ for $x \in U \cap S$ to the right of p and $f(x) < f(p)$ for $x \in U \cap S$ to the left of p:

$$\begin{aligned} &f(x) > f(p) \text{ for all } x \in U \cap S \text{ for which } x > p \\ &f(x) < f(p) \text{ for all } x \in U \cap S \text{ for which } x < p \end{aligned} \tag{23.1}$$

Thus, roughly put, if the slope of $y = f(x)$ is > 0 at a point p then just to the *right* of p the values of f are *higher* than $f(p)$ and just to the *left* of p the values of f are *lower* than $f(p)$.

Proof. Recall the definition of $f'(p)$:

$$f'(p) = \lim_{x\to p} \frac{f(x) - f(p)}{x - p}.$$

If this is > 0 then the ratio

$$\frac{f(x) - f(p)}{x - p}$$

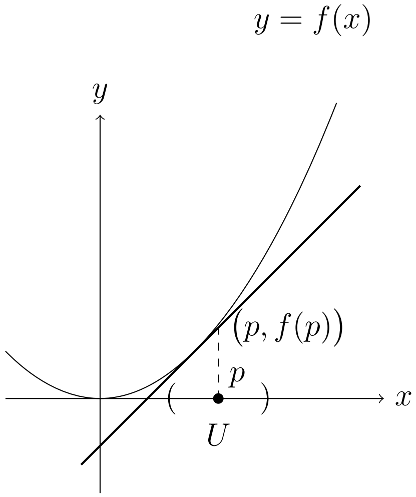


Figure 23.1: Positive slope and increasing function.

is also > 0 when x is near p, but $\neq p$ (see Proposition 10.1.1 for this). Put more precisely, this means that there is a neighborhood U of p such that

$$\tfrac{f(x)-f(p)}{x-p} > 0 \text{ for all } x \in U \cap S \text{ with } x \neq p.$$

If we take an $x \in U \cap S$ to the right of p, we have $x - p > 0$ and so the

$$f(x) - f(p) = (x - p)\left[\frac{f(x) - f(p)}{x - p}\right] > 0.$$

This means $f(x) > f(p)$ for such values of x.

On the other hand, if $x \in U \cap S$ is to the left of p, we have $x - p < 0$ and so the

$$f(x) - f(p) = (x - p)\left[\frac{f(x) - f(p)}{x - p}\right] < 0.$$

This means $f(x) < f(p)$ for such values of x. $\boxed{\text{QED}}$

Using this we can step up to another result going in the converse direction to Proposition 23.1.1:

Proposition 23.1.3 *Suppose f is a continuous function defined on an interval U, and suppose $f'(p)$ exists and is positive (this means > 0) for all p in the interior of U. Then f is* strictly increasing *on U in the sense that:*

$$f(x_1) < f(x_2) \qquad \textit{for all } x_1, x_2 \in U \textit{ with } x_1 < x_2.$$

If f' is assumed to be ≥ 0 on U then the conclusion is $f(x_1) \leq f(x_2)$.

Proof. Consider $x_1, x_2 \in U$ with $x_1 < x_2$. By the mean value theorem there is a point $c \in (x_1, x_2)$ where the derivative $f'(c)$ is given by

$$f'(c) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}.$$

We are given that $f'(c) > 0$ and we know that the denominator $x_2 - x_1$ is positive; hence so is the numerator:

$$f(x_2) - f(x_1) = (x_2 - x_1)\frac{f(x_2) - f(x_1)}{x_2 - x_1} = (x_2 - x_1)f'(c) > 0,$$

which shows that

$$f(x_2) > f(x_1).$$

If we assume only that $f' \geq 0$ then the same argument shows that $f(x_2) \geq f(x_1)$. $\boxed{\text{QED}}$

Here is a slight but useful sharpening of the preceding result:

Proposition 23.1.4 *Suppose f is a continuous function defined on an interval U, and suppose $f'(p)$ exists and is ≥ 0 for all p in the interior of U and $f'(p)$ is 0 at most at finitely many $p \in U$. Then f is strictly increasing on U. If $f' \leq 0$ in the interior of U and $f'(p)$ is 0 at finitely many points $p \in U$ then f is strictly decreasing on U.*

Proof. Assume $f' \geq 0$ in the interior of U and f' takes the value 0 at finitely many points. By Proposition 23.1.3 f is an increasing function on U in the sense that $f(a) \leq f(b)$ for all $a, b \in U$ with $a \leq b$. Hence if $f(s) = f(t)$ for some $s, t \in U$ with $s < t$ then f would be *constant* on the interval $[s, t]$ which would imply that f' is 0 on this entire interval, contradicting the assumption on f'. This proves the result for $f' \geq 0$. For $f' \leq 0$ the argument is exactly similar (or observe that it follows from the case $f' \geq 0$ by flipping the sign of f to $-f$). $\boxed{\text{QED}}$

23.2 Negative derivative and decreasing nature

The results of the preceding section can be run analogously for functions with downward pointing slope.

Proposition 23.2.1 *Let f be a function on a set $S \subset \mathbb{R}$, and p a point in S where $f'(p)$ exists and is negative, that is*

$$f'(p) < 0.$$

Then there is a neighborhood U of p such that the $f(x) < f(p)$ for $x \in U \cap S$ to the right of p and $f(x) > f(p)$ for $x \in U \cap S$ to the left of p:

$$\begin{aligned} &f(x) < f(p) \textit{ for all } x \in U \cap S \textit{ for which } x > p \\ &f(x) > f(p) \textit{ for all } x \in U \cap S \textit{ for which } x < p \end{aligned} \tag{23.2}$$

If f slopes downward along an interval then it is decreasing:

Proposition 23.2.2 *If f is defined on an interval $[a, b]$, where $a, b \in \mathbb{R}$ with $a < b$, and if $f'(p)$ exists and is negaive, that is < 0, for all $p \in [a, b]$ then f is* strictly decreasing *on $[a, b]$ in the sense that:*

$$f(x_1) > f(x_2) \qquad \textit{for all } x_1, x_2 \in [a, b] \textit{ with } x_1 < x_2.$$

If f' is assumed to be ≤ 0 on $[a, b]$ then the conclusion is $f(x_1) \geq f(x_2)$.

23.3 Zero slope and constant functions

Clearly a constant function has zero slope: the derivative of a constant function is 0 wherever defined. One can run this also in the converse direction, but with just a bit of care.

Consider a function G that is defined on a domain consisting of two separated intervals, on each of which it is constant:

$$G(x) = \begin{cases} 1 & \text{if } x \in (0, 1); \\ 4 & \text{if } x \in (8, 9). \end{cases}$$

Then clearly

$$G'(p) = 0 \quad \text{for all } p \text{ in the domain of } G,$$

and yet G is, of course, not constant. On the other hand it is also clear that G really is constant, separately on each interval on which it is defined.

Proposition 23.3.1 *Suppose f is a function on an interval $[a, b]$, where $a, b \in \mathbb{R}$ with $a < b$, and $f'(p) = 0$ for all $p \in [a, b]$. Then f is constant on $[a, b]$. If f is defined on an open interval (a, b) and f' is 0 on (a, b) then f is constant on (a, b).*

One can tinker with this as usual. It is not necessary (for the case of $[a, b]$) to assume that $f'(a)$ and $f'(b)$ to exist; it suffices to assume that f is continuous at a and at b.
Proof. Consider any $x_1, x_2 \in [a, b]$ with $x_1 < x_2$. Then by the mean value theorem

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} = f'(c),$$

for some $c \in (x_1, x_2)$. So if f' is 0 everywhere it follows that

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} = 0,$$

and so

$$f(x_2) - f(x_1) = 0,$$

which means $f(x_1) = f(x_2)$. Thus the values of f at any two different points are equal; that is, f is constant. $\boxed{\text{QED}}$

Chapter 24

Differentiating Inverse Functions

Often an equation of the form

$$y = f(x)$$

can be solved for x:

$$x = f^{-1}(y),$$

and f^{-1} is called an *inverse* to the function f. If f is differentiable then formal common sense suggests that the derivative of the inverse function should be

$$\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}} = \frac{1}{f'(x)}.$$

For example, for

$$y = x^2$$

we have an inverse function given by the square root function

$$x = \sqrt{y}$$

and its derivative should be

$$\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}} = \frac{1}{2x} = \frac{1}{2\sqrt{y}},$$

which certainly is the derivative of $\sqrt{y}$ with respect to y. Of course, we need to avoid the points where dy/dx is 0 (or undefined).

Note that we have been referring to ‘an’ inverse function. For $y = x^2$ another choice of inverse is given by the other ‘branch’ of square root:

$$x = -\sqrt{y}.$$

Things could be really made messy by choosing an inverse function that switches wildly back and forth between the branches $\sqrt{y}$ and $-\sqrt{y}$. This just means that we need to exercise some care about choosing a specific well-behaved branch as an inverse functions.

24.1 Inverses and Derivatives

Suppose f is a function on an interval U such that $f'(x)$ exists for every $x \in U$ and is positive, that is $f'(x) > 0$ (alternatively we could assume that $f' < 0$ everywhere on U). Let V denote the *range* of f:

$$V = f(U) = \{f(x) \, : x \in U\}.$$

Since $f' > 0$ on U, f is a strictly increasing function and so it has a unique inverse function

$$f^{-1} : V \to \mathbb{R},$$

specified by the requirement that

$$f\bigl(f^{-1}(y)\bigr) = y \qquad \text{for all } y \in V.$$

Alternatively,

$$f^{-1}\bigl(f(x)\bigr) = x \qquad \text{for all } x \in U.$$

Proposition 24.1.1 *Suppose f is a function defined on an interval U, such that $f'(x)$ exists and is ≥ 0 for all $x \in U$, being equal to 0 at most at finitely many points. Then $\bigl(f^{-1}\bigr)'(y)$ exists for all $y \in V$, the range of f, and*

$$\bigl(f^{-1}\bigr)'(y) = \frac{1}{f'(x)} \tag{24.1}$$

where $x = f^{-1}(y)$; in (26.39) we take the right side $1/f'(x)$ to be 0 in case $f'(x)$ is ∞, and ∞ if $f'(x)$ is 0.

Note that, as consequence, if f is differentiable (finite derivative) and has derivative either positive on all of U or negative on all of U then f^{-1} is also differentiable. As usual, there is a corresponding result if f' is ≤ 0, being < 0 at all but finitely many points.

Proof. Suppose $f' \geq 0$ on U and is actually > 0 except possibly at finitely many points. Then by Proposition 23.1.4, f is strictly increasing.

By Proposition 11.3.2 the range V of f is an interval and the inverse function f^{-1} is defined on V. Let $p \in U$ and $q = f(p) \in V$. Then

$$(f^{-1})'(q) = \lim_{y \to q} \frac{f^{-1}(y) - f^{-1}(q)}{y - q}. \tag{24.2}$$

Writing x for $f^{-1}(y)$ and p for $f^{-1}(q)$ we see that the difference ratio on the right here is

$$\frac{x - p}{f(x) - f(p)}.$$

Let us denote this by $D(x)$:

$$D(x) = \frac{x - p}{f(x) - f(p)}$$

for $x \in U$ and $x \neq p$ (note that then $f(x) \neq f(p)$, being either $>$ or $<$ than $f(p)$ depending on whether $x > p$ or $x < p$). Thus (24.2) reads:

$$(f^{-1})'(q) = \lim_{y \to q} D\big(f^{-1}(y)\big). \tag{24.3}$$

The definition of $f'(p)$ implies that

$$\lim_{x \to p} D(x) = \lim_{x \to p} \frac{1}{\frac{f(x) - f(p)}{x - p}} = \frac{1}{f'(p)}, \tag{24.4}$$

this being taken to be 0 if $f'(p) = \infty$ and to be ∞ if $f'(p)$ is 0 (these extreme cases require some care).

Now since f^{-1} is continuous (by Proposition 11.3.2) we have:

$$f^{-1}(y) \to p \text{ as } y \to q,$$

and we also know that $f^{-1}(y) \neq p$ when $y \neq p$. Then by Proposition 7.5.1 we have

$$\lim_{y \to q} D\big(f^{-1}(y)\big) = \lim_{x \to p} D(x). \tag{24.5}$$

Combining this with the expression for $(f^{-1})'(q)$ in (24.3) and the limit value in (24.4) we have:

$$(f^{-1})'(q) = \frac{1}{f'(p)}. \tag{24.6}$$

This completes the proof. $\boxed{\text{QED}}$

Chapter 25

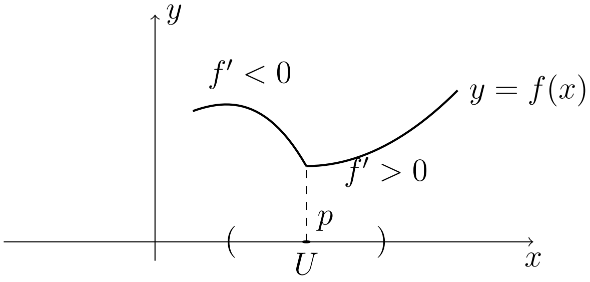
Analyzing local extrema with higher derivatives

25.1 Local extrema and slope behavior

Suppose a function f is defined on a neighborhood of a point $p \in \mathbb{R}$ and f has a local minimum or maximum at p. As we have seen, if $f'(p)$ exists it must be 0. In this secton we explore ways to tell whether p is a local minimum or a local maximum by observing the behavior of the slope f'.

The basic idea is simple: if a continuous graph is sloping downward just to the left of a point p and sloping upward just to the right then it must have a local minimum at p. This is formalized in the following result.

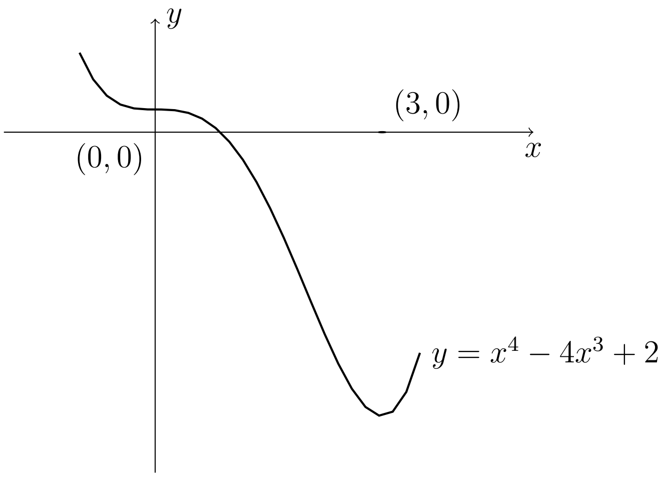
Proposition 25.1.1 *Suppose f is defined and continuous on a neighborhood U of $p \in \mathbb{R}$, and the derivative f' is ≥ 0 to the right of p and ≤ 0 to the left of p; more precisely, suppose f is differentiable on U except possibly at p, and $f'(x) \geq 0$ for $x \in U$ with $x > p$ and $f'(x) \leq 0$ for $x \in U$ with $x < p$.*



Then p is a local minimum for f.

As an example, consider the function

$$g(x) = x^4 - 4x^3 + 2$$



Its derivative is

$$g'(x) = 4x^3 - 12x^2 = 4x^2(x-3),$$

and this is 0 at $x = 0$ and at $x = 3$:

$$g'(0) = 0 \quad \text{and} \quad g'(3) = 0.$$

Let us look at the point $x = 3$. We see that

$$\begin{aligned} 4x^2(x-3) &< 0 \quad \text{if } x < 3; \\ &> 0 \quad \text{if } x > 3. \end{aligned}$$

Then Proposition 25.1.1 implies that $x = 3$ gives a local minimum for g.

What about the point $x = 0$? Observe that on the neighborhood $(-3, 3)$ of 0 we have

$$4x^2(x-3) \leq 0 \text{ for all } x \in (-3, 3).$$

Thus $g(x)$ continues to decrease in value as x passes from the left of 0 to the right of 0, and 0 is *not* a local maximum or minimum.

Proof of Proposition 25.1.1 . Consider any $x \in U$ to the right of p, that is $x > p$; the mean value theorem (Theorem 22.2.1) says that

$$f(x) - f(p) = (x - p)f'(c) \quad \text{for some } c \in (p, x).$$

If f' is ≥ 0 on U to the right of p then $f'(c) \geq 0$, and so we see that

$$f(x) - f(p) \geq 0 \quad \text{for } x \in U, \text{ with } x > p.$$

Thus

$$f(x) \geq f(p) \quad \text{for } x \in U, \text{ with } x > p.$$

On the other hand, taking $x < p$ but inside U we have, again by the mean value theorem,

$$f(x) - f(p) = (x - p)f'(c) \quad \text{for some } c \in (x, p),$$

but observe now that $x - p < 0$ and $f'(c)$ is given to be ≥ 0 (for c is to the left of p); hence

$$f(x) - f(p) \geq 0 \quad \text{for } x \in U, \text{ with } x < p.$$

Thus,

$$f(x) \geq f(p) \quad \text{for } x \in U, \text{ with } x < p.$$

We have shown that $f(x)$ is $\geq f(p)$ for all $x \in U$, both those to the left of p and those to the right of p. This means that f has a local minimum at p. $\boxed{\text{QED}}$

By a closely similar argument we obtain the analogous result for local maxima:

Proposition 25.1.2 *Suppose f is defined and continuous on a neighborhood U of $p \in \mathbb{R}$, and the derivative f' is ≤ 0 to the right of p and ≥ 0 to the left of p; more precisely, suppose f is differentiable on U except possibly at p, and $f'(x) \leq 0$ for $x \in U$ with $x > p$ and $f'(x) \geq 0$ for $x \in U$ with $x < p$. Then p is a local maximum for f.*

25.2 The second derivative test

There is often a faster way to check whether a point p where $f'(p) = 0$ is a local minimum of maximum: this is by simply computing the *second derivative* $f''(p)$. If this is positive we have a local minimum at p, while if it is negative then we have a local maximum at p.

For example,

$$y = \sin(x^2)$$

has derivative

$$\frac{d\sin(x^2)}{dx} = 2x\cos(x^2),$$

which is 0 when $x = 0$, and second derivative

$$\frac{d^2\sin(x^2)}{dx^2} = 2\cos(x^2) - 4x^2\sin(x^2)$$

whose value at $x = 0$ is 2; then, without even knowing anything about the graph, we can say that $\sin x^2$ has a local minimum at $x = 0$. (You should, however, see that since $\sin\theta \simeq \theta$ near $\theta = 0$, the graph $y = \sin(x^2)$ looks about like that of $y = x^2$, near $x = 0$, and since this clearly has a local minimum at $x = 0$, it is a good guess that $\sin(x^2)$ has a local minimum at $x = 0$.)

Here is a formal statement:

Proposition 25.2.1 *Let f be a function defined and differentiable on a neighborhood of $p \in \mathbb{R}$. Assume also that the second derivative $f''(p)$ exists. If*

$$f'(p) = 0 \qquad \textit{and} \qquad f''(p) > 0$$

then f has a local minimum at p. If

$$f'(p) = 0 \qquad \textit{and} \qquad f''(p) < 0$$

then f has a local maximum at p.

For the proof take a look back first at Propositions 23.1.2 and **??**.

Proof. We are given that f is differentiable on the neighborhood W of p. Thus the derivative $f'(x)$ exists and is finite at all $x \in W$. Moreover, we are also given that f' itself has a derivative $(f')'(p)$ at the point p.

Suppose $f''(p) > 0$. Thus, the derivative of f' at p is > 0. Then by Proposition 23.1.2, the value $f'(x)$ for x immediately to the left of p is less than the value $f'(p)$, whereas the value $f'(x)$ for x immediately to the right of p is greater than the value $f'(p)$. More precisely, there is a neighborhood U of p such that

$$\begin{aligned} f'(x) &> f'(p) \quad \text{for } x > p \text{ and } x \in U; \\ &< f'(p) \quad \text{for } x < p \text{ and } x \in U. \end{aligned}$$

Since we are given that $f'(p)$ is 0 this means

$$\begin{aligned} f'(x) &> 0) \quad \text{for } x > p \text{ and } x \in U; \\ &< 0 \quad \text{for } x < p \text{ and } x \in U. \end{aligned}$$

Then by Proposition 25.1.1, f has a local minimum at p.

The argument for $f''(p) < 0$ is very similar, using Proposition **??** to see first that f' is positive to the left of p and negative to the right of p and concluding then that p is a loca maximum for f. $\boxed{\text{QED}}$

There are functions for which both first and second derivatives are 0 at the same point, and then we cannot draw any conclusions about local maximum/minimum at that point. For example,

$$y = x^4$$

is 'very flat' at $x = 0$ since both its first derivative, which is $4x^3$, and its second derivative $12x^3$ are 0 at $x = 0$. The point $x = 0$ is in fact a local minimum for x^4 but we cannot see this simply using the second derivative test.

As another example of what can go wrong, consider

$$g(x) = x^4 - 4x^3 + 2$$

The derivative is

$$g'(x) = 4x^3 - 12x^2 = 4x^2(x - 3)$$

and the second derivative is

$$g''(x) = 12x^2 - 24x = 12x(x - 2).$$

Thus

$$g'(0) = 0, \quad \text{and} \quad g'(3) = 0$$

and

$$g''(0) = 0, \quad \text{and} \quad g''(3) = 36 > 0.$$

Thus $x = 3$ gives a local minimum, but we cannot draw any conclusions about $x = 0$ from the second derivative test.

Chapter 26

Exp and Log

The *exponential function* is one of the most useful functions in mathematics, and is expressed through the amazing formula

$$\left[1+1+\frac{1}{2!}+\frac{1}{3!}+\cdots\right]^x = 1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots\,. \tag{26.1}$$

Its inverse function is the *natural logarithm* log. In this chapter we study the basic properties of these two functions.

The history of the discovery/understanding of these two functions is an entertaining example of how mathematical concepts develop trough unexpected twists and turns and near-misses [1, 2, 4]. Our approach is not historical; we first summarize the essential facts about e^x and $\log(x)$ that explain how to work with these functions, and then we give a logical development of the theory. This approach is fast but gives little insight on how or why these ideas were developed historically.

26.1 Exp summarized

The function exp is defined on $\mathbb{R}$ by

$$\exp(x) = 1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots$$

for all $x \in \mathbb{R}$. It is a fact that this is a real number (finite) for all $x \in \mathbb{R}$.

The number $\exp(1)$ is denoted e:

$$e = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \simeq 2.718281\ldots \tag{26.2}$$

The importance of e lies in the amazing identity

$$\exp(x) = e^x,$$

which is another way of writing (26.1). What exactly it means to raise e to the power x, which is a real number, will be explored more carefully later.

The value of exp at 0 is clearly equal to 1:

$$\exp(0) = 1.$$

Moreover, the derivative of exp is again exp:

$$\exp' = \exp, \tag{26.3}$$

which can also be expressed as

$$\frac{de^x}{dx} = e^x. \tag{26.4}$$

This property along with the value at 0 *uniquely characterizes* the function exp: any function on $\mathbb{R}$ whose derivative is itself and whose value at 0 is 1 is the exponential function.

Figure 26.1 shows the graph of $y = e^x$.

The graph of the exponential function shows rapid increase as $x \to \infty$ and rapid decay to 0 as $x \to -\infty$:

$$\begin{aligned} \lim_{x\to\infty} e^x &= \infty \\ \lim_{x\to-\infty} e^x &= 0. \end{aligned} \tag{26.5}$$

Among the early studies of the exponential function is its use in describing the growth of money under compound interest. This approach leads to the following limit formulas:

$$e^x = \lim_{n\to\infty}\left(1 + \frac{1}{n}\right)^{nx} = \lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^n. \tag{26.6}$$

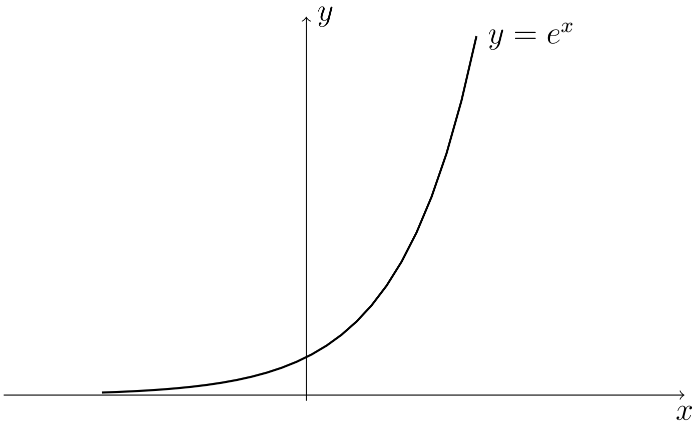


Figure 26.1: Graph of the exponential function.

26.2 Log summarized

The natural *logarithm* is the inverse function of exp: thus $\log(A)$ is the real number with the property that

$$\exp\left(\log(A)\right) = A,$$

or, equivalently,

$$e^{\log A} = A. \tag{26.7}$$

It is defined for all $A > 0$.

For example, since $e^0 = 1$ we have

$$\log 1 = 0.$$

Since $e^1 = 1$ we have

$$\log 1 = e.$$

The notation ln is also used to denote the logarithm.

An alternative way to state the fact that log is inverse to the function exp is

$$\log e^x = x \quad \text{for all } x \in \mathbb{R}. \tag{26.8}$$

The basic algebraic properties of log are:

$$\begin{aligned} \log(AB) &= \log(A) + \log(B) \\ \log(A/B) &= \log(A) - \log(B) \\ \log(A^k) &= k\log(A), \end{aligned} \tag{26.9}$$

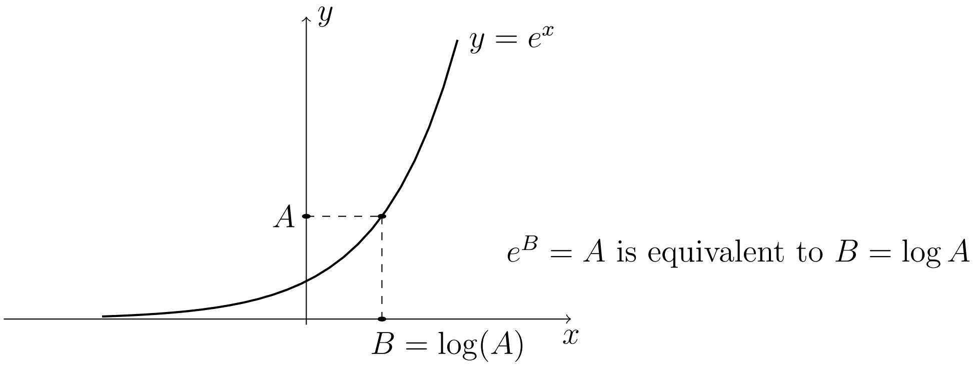


Figure 26.2: $\log(A)$ read off from the graph of $y = e^x$.

for all $A, B > 0$ and $k \in \mathbb{R}$. These properties made log enormously useful in carrying out complex calculations involving multiplication and division. In fact log (up to a scaling) was studied and used well before the exponential function was identified.

The graph of log is the graph of Exp viewed from one side (thus, with x and y axes interchanged). We have the following limits of interest:

$$\begin{aligned} \lim_{x\to\infty} \log(x) &= \infty \\ \lim_{x\to-\infty} \log(x) &= 0. \end{aligned} \tag{26.10}$$

The derivative of log is the reciprocal:

$$\frac{d\log(x)}{dx} = \frac{1}{x} \tag{26.11}$$

for $x > 0$.

Figure 26.3 shows the graph of log:

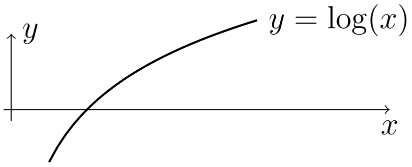


Figure 26.3: The graph of log.

Here are a couple of useful values of log:

$$\log 1 = 0, \qquad \log(e) = 1. \tag{26.12}$$

The function log is strictly increasing, and so

$$\begin{aligned} \log(x) &< 0 \quad \text{if } x < 1; \\ \log(x) &> 0 \quad \text{if } x > 1. \end{aligned} \tag{26.13}$$

26.3 Real Powers

For any real number y we have the positive integer powers

$$y^1 = 1,\ y^2 = y \cdot y,\ y^3 = y^2 \cdot y, \ldots, y^{n+1} = y^n \cdot y.$$

The 0-th power is

$$y^0 = 1.$$

Negative powers are given by reciprocals

$$y^{-n} = \frac{1}{y^n},$$

for any positive integer n. Of course, this makes sense only for $y \neq 0$.

Next we have rational powers. For example,

$$y^{1/2} = \sqrt{y}$$

is defined to be the unique real number ≥ 0 whose square is y:

$$(y^{1/2})^2 = y.$$

Thus $y^{1/2}$ is defined only for $y \geq 0$ (for otherwise we couldn't square something to end up with y.)

More generally, for any positive real number $A > 0$ and integers p and q, with $q \neq 0$, the power

$$A^{p/q}$$

is defined to be the unique positive real number whose q-th power is A^p:

$$(A^q)^{p/q} = A^p.$$

Thus, we have a definition for

$$A^r$$

for all positive real A and all rational r. These definitions are designed to ensure that the following convenient algebra holds:

$$\begin{aligned} A^{r+s} &= A^r A^s \\ (A^r)^s &= A^{rs} \\ (AB)^r &= A^r B^r \end{aligned} \tag{26.14}$$

for all positive real A, B and rationals r and s.

Moving on to real powers, it is natural to define

$$A^x = \lim_{r \to x, q \in \mathbb{Q}} A^r. \tag{26.15}$$

That this exists for all positive real A and $x \in \mathbb{R}$ is intuitively clear but not simple to prove. Perhaps the shortest way to see that A^x exists is by using log: for rational r we have

$$A^r = \left(e^{\log(A)}\right)^r = e^{(\log(A))r} = e^{r\log(A)} = \exp\left(r\log(A)\right).$$

Since the function exp is continuous (it is, in fact, differentiable), we can take then limit $r \to x$, for any real number x, to obtain:

$$A^x = \exp\left(x\log(A)\right). \tag{26.16}$$

Taking $A = e$ confirms the expected result

$$e^x = \exp(x) \quad \text{for all } x \in \mathbb{R}.$$

We can also verify the algebraic relations:

$$\begin{aligned} A^{x+y} &= A^x A^y \\ (A^x)^y &= A^{xy} \\ (AB)^x &= A^x B^x, \end{aligned} \tag{26.17}$$

for all $A, B, x, y \in \mathbb{R}$ with $A, B > 0$.

Using the derivative of the exponential function we obtain immediately that

$$\frac{dA^x}{dx} = e^{x\log A} \cdot \log(A). \tag{26.18}$$

Note that on the right we have $e^{x\log A}$ which is, in fact, A^x. Thus,

$$\frac{dA^x}{dx} = A^x \log(A).$$

26.4 Example Calculations

To see how to work with derivatives of log and exp we work out a few examples.

We have already observed in (26.18) that

$$\frac{dA^x}{dx} = A^x \log(A), \tag{26.19}$$

for any positive real constant A.

Thus, for example,

$$\begin{aligned} \frac{d2^{(2^x)}}{dx} &= 2^{(2^x)} \log(2) \cdot \frac{d2^x}{dx} \qquad \text{(by the chain rule)} \\ &= 2^{(2^x)} \log(2) \cdot 2^x \log(2) \\ &= 2^{2^x+x} (\log 2)^2. \end{aligned}$$

Next consider the derivative of the function x^x. To work this out we rewrite x as $e^{\log(x)}$:

$$x^x = \left(e^{\log(x)}\right)^x = e^{x\log(x)}.$$

Now we can differentiate this:

$$\begin{aligned} \frac{dx^x}{dx} &= \frac{de^{x\log(x)}}{dx} \\ &= e^{x\log(x)} \frac{dx\log(x)}{dx} \qquad \text{(by the chain rule)} \\ &= e^{x\log(x)} \left[1 \cdot \log(x) + x \cdot \frac{1}{x}\right] \\ &= x^x \left[\log x + 1\right] \qquad \text{(recognizing } e^{x\log(x)} \text{ as } x^x\text{).} \end{aligned}$$

Note that

$$1 + \log(x) = \log e + \log(x) = log(ex).$$

Similarly,

$$\frac{dx^{1/x}}{dx} = x^{1/x} \left[\frac{1 - \log x}{x^2}\right]. \tag{26.20}$$

Observe that this is > 0 when $x < e$ and is < 0 when $x > e$. This means that $x^{1/x}$ is increasing for $x < e$ and decreasing for $x < 0$. So $x^{1/x}$ is maximum at e:

$$\max_{x>0} x^{1/x} = e^{1/e}.$$

26.5 Proofs for Exp and Log

To really develop the theory and results for exp and log we should start with the definition of $\exp(x)$ as

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

However, we have not yet developed enough working tools for such *power series*, and so we will follow a modest strategy here. We *assume* that there exists a function exp defined on $\mathbb{R}$ that is differentiable and satisfies the following conditions:

$$\begin{aligned} \exp' &= \exp \\ \exp(0) &= 1. \end{aligned} \tag{26.21}$$

We will prove that there can be at most one such function and then prove the crucial relation

$$\exp(x) = e^x, \quad \text{where } e = \exp(1).$$

Observe by the chain rule that

$$\frac{d\exp(Kx)}{dx} = \exp(Kx) \cdot K = K\exp(Kx),$$

for any $K \in \mathbb{R}$. In particular,

$$\frac{d\exp(-x)}{dx} = -\exp(-x).$$

Proposition 26.5.1 *For any $x \in \mathbb{R}$ the value $\exp(x)$ is not 0, and, moreover,*

$$\exp(x)\exp(-x) = 1. \tag{26.22}$$

Note that this says

$$\exp(-x) = \frac{1}{\exp(x)}. \tag{26.23}$$

Proof. First we take the derivative of $\exp(x)\exp(-x)$ and show that it is 0:

$$\begin{aligned} \frac{d\exp(x)\exp(-x)}{dx} &= \frac{d\exp(x)}{dx}\exp(-x) + \exp(x)\frac{d\exp(-x)}{dx} \\ &= \exp(x)\exp(-x) - \exp(x)\exp(-x) \\ &= 0. \end{aligned}$$

Thus $\exp(x)\exp(-x)$ is *constant*, and so

$$\exp(x)\exp(-x) = \exp(0)\exp(-0) = 1 \cdot 1 = 1,$$

for all $x \in \mathbb{R}$. This proves (26.22) and from this relation it is clear that $\exp(x) \neq 0$. $\boxed{\text{QED}}$

The stragegy used in the proof above, showing that the derivative of a function is 0 and then concluding that its value is constant, equal to the value at $x = 0$, will be used several times.

The following result shows that the function exp is uniquely specified by the condition that $\exp' = \exp$ and the 'initial' value $\exp(0) = 1$:

Proposition 26.5.2 *Suppose F is a differentiable function on $\mathbb{R}$ whose derivative is equal to itself:*

$$F' = F.$$

Then F is a constant multiples of exp*:*

$$F(x) = F(0)\exp(x) \quad \textit{for all } x \in \mathbb{R}.$$

Proof. Consider the function $F(x)/\exp(x)$, which is defined for all $x \in \mathbb{R}$ since the denominator $\exp(x)$ is never 0 (Proposition 26.5.1). Taking the derivative we have

$$\begin{aligned} \frac{d\frac{F(x)}{\exp(x)}}{dx} &= \frac{\exp(x)F'(x) - F(x)\exp'(x)}{(\exp(x))^2} \quad \text{(by the quotient rule)} \\ &= \frac{\exp(x)F(x) - F(x)\exp(x)}{(\exp(x))^2} \\ &= 0. \end{aligned}$$

Thus, $F(x)/\exp(x)$ is *constant*, and so

$$\frac{F(x)}{\exp(x)} = \frac{F(0)}{\exp(0)} = \frac{F(0)}{1} = F(0),$$

for all $x \in \mathbb{R}$. Hence

$$F(x) = F(0)\exp(x)$$

for all $x \in \mathbb{R}$. $\boxed{\text{QED}}$

Next we can prove a key algebraic property for exp:

Proposition 26.5.3 *For every $a, b \in \mathbb{R}$ we have*

$$\exp(a+b) = \exp(a)\exp(b). \tag{26.24}$$

Proof. Consider any $a \in \mathbb{R}$ and let F be the function

$$F(x) = \exp(a+x) \qquad \text{for all } x \in \mathbb{R}.$$

Then

$$F'(x) = \exp'(a+x) \cdot 1 = \exp(a+x) = F(x).$$

Then by Proposition 26.5.2 we conclude that

$$F(x) = F(0)\exp(x) \text{ for all } x \in \mathbb{R}.$$

Observing that

$$F(0) = \exp(a),$$

we conclude that

$$F(x) = \exp(a)\exp(x) \text{ for all } x \in \mathbb{R}.$$

Recalling that $F(x)$ is $\exp(a+x)$ we are done. $\boxed{\text{QED}}$

It is useful to observe that this stage that $\exp(x)$ is strictly positive:

Proposition 26.5.4 *The function* exp *assumes only positive values:*

$$\exp(x) > 0 \qquad \textit{for all } x \in \mathbb{R}. \tag{26.25}$$

Proof. This follows from writing x as $x/2 + x/2$ and using the previous Proposition:

$$\exp(x) = \exp\left(\frac{x}{2} + \frac{x}{2}\right) = \exp(x/2)\exp(x/2) = [\exp(x/2)]^2.$$

This, being a square, is ≥ 0. Moreover, we know from Proposition 26.5.1 that $\exp(x/2)$ is not 0. Hence $\exp(x)$ is actually > 0. $\boxed{\text{QED}}$

From the exponential multiplicative property in Proposition 26.5.3 we have

$$\exp(2a) = \exp(a+a) = [\exp(a)]^2$$

and

$$\exp(3a) = \exp(2a+a) = \exp(2a)\exp(a) = [\exp(a)]^2\exp(a) = [\exp(a)]^3.$$

This line if reasoning implies that

$$\exp(Na) = [\exp(a)]^N \tag{26.26}$$

for all $a \in \mathbb{R}$ and $N \in \{1, 2, 3, \ldots\}$. Next, for such a and N we have

$$\begin{aligned}\exp(-Na) &= \frac{1}{\exp(Na)} \quad \text{(by (26.23))}\\ &= \frac{1}{[\exp(a)]^N}\\ &= [\exp(a)]^{-N}.\end{aligned}$$

Thus we have

$$\exp(na) = [\exp(a)]^n \tag{26.27}$$

for $a \in \mathbb{R}$ and *all integers* $n \in \mathbb{Z}$ (you can check the case $n = 0$ directly: both sides are 1 in that case).

To proceed to rational powers first consider a simple case $\exp\left(\frac{1}{2}a\right)$; this is a positive real number whose square is

$$\left[\exp\left(\frac{1}{2}a\right)\right]^2 = \exp\left(2\frac{1}{2}a\right) = \exp(a).$$

Hence $\exp(\frac{1}{2}a)$ is the positive square-root of $\exp(a)$:

$$\exp\left(\frac{1}{2}a\right) = [\exp(a)]^{1/2}.$$

Proceeding on to a general rational number

$$r = \frac{p}{q} \qquad \text{where } p, q \in \mathbb{Z} \text{ and } q \neq 0,$$

we have

$$[\exp(ra)]^q = \exp(qra) = \exp(pa) = [\exp(a)]^p,$$

and so $\exp(ra)$ is a positive real number whose q-th power is $[\exp(a)]^p$. Then, by defininition of $(\cdot)^{p/q}$, we have

$$\exp(ra) = [\exp(a)]^{p/q}.$$

Thus,

$$\exp(ra) = [\exp(a)]^r \tag{26.28}$$

for all $a \in \mathbb{R}$ and all $r \in \mathbb{Q}$.

Finally, we can proceed to an arbitary real power. Consider any $a, x \in \mathbb{R}$. The definition of the x-th power is that

$$[\exp(a)]^x = \lim_{r \to x, r \in \mathbb{Q}} [\exp(a)]^r .$$

But we know that for rational r we have

$$[\exp(a)]^r = \exp(ra),$$

and so, since the differentiable function exp is continuous, we have

$$\lim_{r \to x, r \in \mathbb{Q}} [\exp(a)]^r = \exp(xa).$$

Putting everything together we have:

Proposition 26.5.5 *For any real numbers a and x we have*

$$\exp(ax) = [\exp(a)]^x . \tag{26.29}$$

Now we specialize this to $a = 1$ to obtain the crucial formula:

$$\exp(x) = e^x \qquad \text{for all } x \in \mathbb{R}, \tag{26.30}$$

where e is the value of exp at 1:

$$e \stackrel{\text{def}}{=} \exp(1). \tag{26.31}$$

The derivative of exp is exp, by assumption in our approach, and this can now be displayed as

$$\frac{de^x}{dx} = e^x.$$

Proposition 26.5.6 *The exponential function is strictly increasing:*

$$e^a < e^b \qquad \textit{for all } a, b \in \mathbb{R} \textit{ with } a < b. \tag{26.32}$$

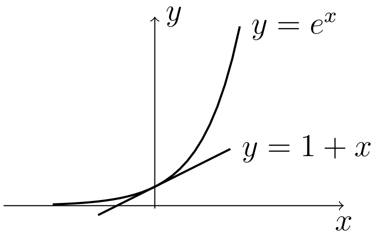


Figure 26.4: The inequality $e^x \geq 1 + x$ in terms of the tangent at $(0, 1)$.

Proof. The derivative of exp is exp, and this is always positive. Hence exp is strictly increasing. $\boxed{\text{QED}}$

Taking one of the numbers a and b to be 0 and using $e^-0 = 1$ we conclude that

$$\begin{aligned} e^x &> 1 \quad \text{if } x > 0; \\ e^x &< 1 \quad \text{if } x < 0. \end{aligned} \tag{26.33}$$

Proposition 26.5.7 *The function* exp *satisfies*

$$\exp(x) \geq 1 + x \qquad \textit{for all } x \in \mathbb{R}, \tag{26.34}$$

with $=$ *holding only when* $x = 0$.

A geometric way of understanding this inequality is that the graph of $y = e^x$ lies above the tangent line $y = x + 1$ at $(0, 1)$ (see Figure 26.4.)

Proof. First observe that

$$\frac{d\left(e^x - (1 + x)\right)}{dx} = e^x - 1.$$

This is positive when $x > 0$ and negative when $x < 0$. Thus, $e^x - (1 + x)$ attains its minmum value when $x = 0$:

$$e^x - (1 + x) > e^0 - (1 + 0) = 1 - 1 = 0 \qquad \text{for all } x > 0 \text{ and all } x < 0.$$

This proves the inequality (26.34). $\boxed{\text{QED}}$

One consequence of the inequality (26.34) is the $\exp(x)$ goes to ∞ when $x \to \infty$:

$$\lim_{x \to \infty} e^x = \infty. \tag{26.35}$$

Working with $e^{-w} = 1/e^w$ we then see that

$$\lim_{x\to-\infty} e^x = 0. \tag{26.36}$$

The intermediate value theorem implies that the function exp has an inverse defined on all positive real numbers; this function is log:

$$\log = \exp^{-1} : (0,\infty) \to \mathbb{R}, \tag{26.37}$$

satisfying

$$\begin{aligned} \exp\left(\log(A)\right) &= A \quad \text{for all } A \in (0,\infty);\\ \log\left(\exp(B)\right) &= B \quad \text{for all } B \in \mathbb{R}. \end{aligned} \tag{26.38}$$

If a function f, defined on an open interval U, has an inverse f^{-1}, and if f is differentiable at $p \in U$ with $f'(p) \neq 0$ then the derivative of the inverse function is given by

$$(f^{-1})'(q) = \frac{1}{f'(p)}, \tag{26.39}$$

where $q = f(p)$, which means $p = f^{-1}(q)$. Applying this to the exponential function and its inverse, log, we have first

$$\log' q = \frac{1}{e^p} = \frac{1}{q},$$

for all $p \in \mathbb{R}$ and $q = e^p$. Restating in different notation this says:

$$\frac{d\log x}{dx} = \frac{1}{x} \qquad \text{for } x \in (0,\infty). \tag{26.40}$$

Chapter 27

Convexity

Convexity is a powerful notion. It is useful in establishing results about maxima and minima, and useful in many different applications. This chapter gives an initial glimpse of this deep and varied subject.

27.1 Convex and concave functions

A function f is said to be *convex* on an interval U if for any points $a, b \in U$, the graph $y = f(x)$, over the interval $x \in [a, b]$, below the secant line segment joining the point $\big(a, f(a)\big)$ with $\big(b, f(b)\big)$.

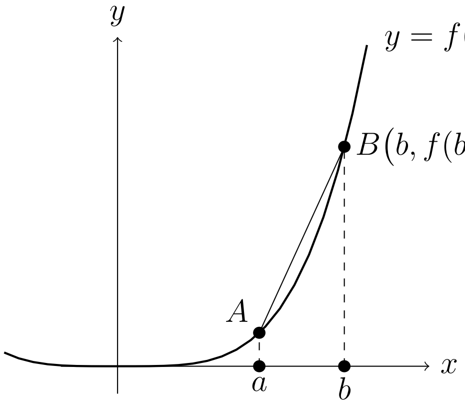


Figure 27.1: Convex function: the graph lies below secant segments

By 'below' we allow the possibility that the graph and the secant segment might touch or run along each other. For example, *a function whose graph is a straight line is convex.*

A function f is *strictly convex* on an interval U if the graph of f between any two points in U lies *strictly below* the corresponding secant segment.

A function f is *concave* over an interval U if the graph of f between any $a, b \in U$ lies above the secant segment joining $A\big(A, f(a)\big)$ and $B\big(b, f(b)\big)$. It is *strictly concave* of the graph lies strictly above all secant segments.

Thus, *a function whose graph is a straight line is both convex and concave.*

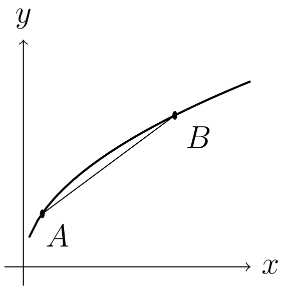


Figure 27.2: Concave function: the graph lies above secant segments

Note that we only consider the notions of convexity and concavity of functions defined on *intervals.*

27.2 Convexity and slope

Consider a function f on an interval U. Let us take three points $a, p, b \in U$ with

$$a < p < b$$

and examine the behavior of the secants over $[a, p]$ and over $[p, b]$.

Let A be the point on the graph of $y = f(x)$ with x-coordinate a; thus A is $\big(a, f(a)\big)$. Next let B be $(b, f(b))$ and P the point $(p, f(p))$. Let Q be the point on the secant segment AB whose x-ccordinate is p. Thus Q is of the form

$$(p, L(p))$$

where $L(p)$ is the y-cordinate of Q.

The condition that the point $P(p, f(p))$ lie below the segment AB means that P is below Q:

$$f(p) \leq L(p).$$

This means that the slope of AP is $\leq$ the slope of AQ. Now the slope of AQ is the same as the slope of AB. Thus the condition that P is below AB is equivalent to

$$\text{slope } AP \leq \text{slope } AB.$$

Similarly, the same condition is also equivalent to

$$\text{slope } AB \leq \text{slope } PB.$$

Thus, convexity of f is equivalent to the relations

$$\text{slope } AP \leq \text{slope } AB \leq \text{slope } PB,$$

for all $a, p, b \in U$, with $p \in [a, b]$.

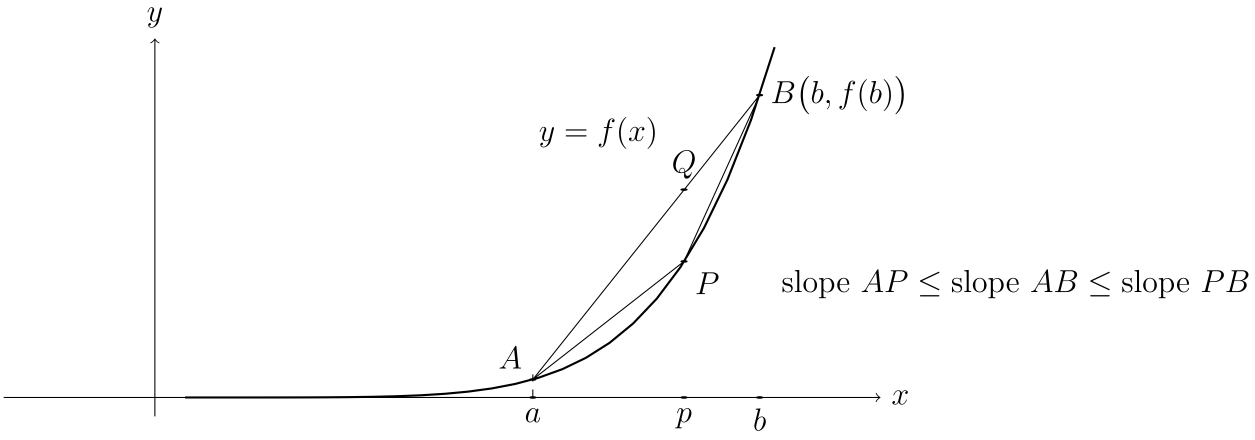


Figure 27.3: Convex function: secant slopes increase

Just the condition

$$\text{slope } AP \leq \text{slope } PB \tag{27.1}$$

forces the slope of AB to lie in between these values. For example, since the slope of AP is $\leq$ the slope of PB, the segment AP produced beyond P would end up below B, and this would mean that the slope of AP is $\leq$ the slope of AB.

Similarly, f is *concave* over an interval if the secant slopes *decrease.*

27.3 Checking convexity/concavity

In this section we will see how to check convexity/concavity of specific functions and how to work with such functions. The methods used will be justified in later sections.

A function f is convex on an interval U if its derivative f' is an increasing function on U (this means $f(s) \leq f(t)$ if $s \leq t$); it is strictly convex if f' is strictly increasing ($f(s) < f(t)$ if $s < t$). Mostly we can check for convexity even more efficiently: if the second derivative f'' is ≥ 0 on U then f is convex on U, and if $f'' > 0$ on U then f is strictly convex on U.

For concavity we have the analogous results. A function f is concave on an interval U if its derivative f' is a decreasing function on U; it is strictly concave if f' is strictly decreasing ($f(s) < f(t)$ if $s < t$). If f'' is ≤ 0 on U then f is concave on U, and if $f'' < 0$ on U then f is strictly concave on U.

Consider the function

$$x^2.$$

Its second derivative is 2:

$$(x^2)' = 2x \quad \text{and} \quad (x^2)'' = 2 > 0,$$

and so x^2 *is strictly convex on* $\mathbb{R}$.

Next consider

$$x^4$$

Its second derivative is $12x^2$, which is ≥ 0 everywhere and so the function is convex. (In fact it is strictly convex even though $12x^2$ hits 0 at the single point $x = 0$.)

More generally, consider

$$x^p,$$

for x in the interval $(0, \infty)$, where p is some constant. The derivative is px^{p-1} and the second derivative is

$$p(p-1)x^{p-2}.$$

If the power p is ≥ 1 then this second derivative is ≥ 0 for all $x \in \mathbb{R}$ and so x^p *is convex on* $(0, \infty)$ *if* $p \geq 1$. In fact, x^p *is strictly convex on* $(0, \infty)$ *if* $p > 1$.

Another example is

$$\frac{1}{x}$$

Its derivative is

$$-x^{-2}$$

and the second derivative is

$$(-1)(2)x^{-3} = 2x^{-3} = \frac{2}{x^3},$$

which is > 0 when $x > 0$ and is < 0 when $x < 0$. Thus

$$1/x \text{ is strictly convex for } x \in (0, \infty),$$

and is strictly concave for $x < 0$.

Now let us look at

$$e^x.$$

The second derivative is e^x, which is positive. Hence, e^x *is strictly convex on* $\mathbb{R}$.

Lastly consider

$$\log x$$

for $x > 0$. The derivative is

$$1/x,$$

which is strictly decreasing. Hence $\log x$ *is strictly concave.*

27.4 Inequalities from convexity/concavity

Recall that a function Φ, on an ineral U, is convex if its graph lies below the secant segments, with strict convexity meaning that the graph always lies strictly below the secant segments. From this one can show that Φ *is convex if and only if*

$$\Phi(\text{weighted average}) \leq \text{weighted average of } \Phi, \tag{27.2}$$

so that, for instance

$$\Phi\left(\frac{1}{3}a + \frac{2}{3}b\right) \leq \frac{1}{3}\Phi(a) + \frac{2}{3}\Phi(b),$$

for all $a, b \in U$. Strict convexity means that here $\leq$ would be replaced by $<$ except in the trivial case where $a = b$. For concavity condition (27.2) is altered by replacing $\leq$ by $\geq$.

The weighted average might involve several points/numbers drawn from U. For example, if Φ is convex on an interval U then

$$\Phi\left(\frac{1}{10}a + \frac{1}{10}b + \frac{8}{10}c\right) \leq \frac{1}{10}\Phi(a) + \frac{1}{10}\Phi(b) + \frac{8}{10}\Phi(c),$$

for all $a, b, c \in U$.

A weighted average is also called a *convex combination.* Thus,

$$\frac{2}{7}a + \frac{1}{7}b + \frac{3}{7}c + \frac{1}{7}d$$

is a convex combination of a, b, c, and d. Thus, a *convex combination of* $p_1, \ldots, p_N \in \mathbb{R}$ *is of the form*

$$w_1p_1 + \cdots + w_Np_N,$$

where the weights $w_1, \ldots, w_N$ *lie in* $[0, 1]$ *and add up to* 1*:*

$$w_1 + \cdots + w_N = 1.$$

Thus, (27.2), written out more formally says that a function Φ on an interval U is convex if and only if

$$\Phi(w_1p_1 + \cdots + w_Np_N) \leq w_1\Phi(p_1) + \cdots + w_N\Phi(p_N) \tag{27.3}$$

for all $p_1, \ldots, p_N \in U$ and all weights $w_1, \ldots, w_N \in [0, 1]$ (adding to 1), for every choice of $N \in \{1, 2, 3, \ldots\}$. The function Φ is *strictly* convex if (27.3) holds with $\leq$ replaced by $<$ if the points $p_1, \ldots, p_N$ are all distinct and none of the weights is 1. For concavity we simply reverse the inequalities.

Let us apply the characterization of convexity given in the inequality (27.3) to the functions we looked at in the preceding section.

For the convex function x^2 we have, using the simplest interesting weighted average:

$$\left(\frac{1}{2}a + \frac{1}{2}b\right)^2 \leq \frac{1}{2}a^2 + \frac{1}{2}b^2. \tag{27.4}$$

Thus, the *the square of an average is at most the average of the squares.* Note that since x^2 is actually strictly convex the inequality $\leq$ can be replaced by $<$, unless $a = b$. With three points we have

$$\left(\frac{1}{3}a + \frac{1}{3}b + \frac{1}{3}c\right)^2 \leq \frac{1}{3}a^2 + \frac{1}{3}b^2\frac{1}{3}c^2. \tag{27.5}$$

More generally, for any $x_1, \ldots, x_N \in \mathbb{R}$ we have

$$\left(\frac{x_1 + \cdots + x_N}{N}\right)^2 \leq \frac{1}{N}\left(x_1^2 + \cdots + x_N^2\right). \tag{27.6}$$

Here the $\leq$ can be replaced by $<$ except in the case all the x_i's are equal. We can also write the inequality (27.6) as:

$$x_1^2 + \cdots + x_N^2 \geq \frac{(x_1 + \cdots + x_N)^2}{N}, \tag{27.7}$$

with equality holding if and only if all the x_i are equal.

Here is a quick application: suppose a length L of wire is cut into N pieces, of lengths $x_1, \ldots, x_N$, and each piece is bent into a square; what should the lengths x_i be in order fot the squares to cover a minimum total area? To answer this notice that the sum of the areas of the squares is

$$\left(\frac{x_1}{4}\right)^2 + \cdots + \left(\frac{x_N}{4}\right)^2 = \frac{x_1^2 + \cdots + x_N^2}{16},$$

and from (27.7) we see that this is

$$\geq \frac{(x_1 + \cdots + x_N)^2/N}{16} = \frac{L^2}{16N},$$

with $\geq$ being $=$ if and only if all the x_i are equal. Thus the minimum area is obtained if the wire is cut into equal lengths and the minimum area thus obtained is $L^2/(16N)$.

Now we turn to another example, the strictly *concave* function log, defined on $(0, \infty)$. Working with the equally weighted average of $a, b > 0$ we have

$$\log\left(\frac{1}{2}a + \frac{1}{2}b\right) \geq \frac{1}{2}\log a + \frac{1}{2}\log b. \tag{27.8}$$

The right side simplifies to

$$\frac{1}{2}(\log a + \log b) = \frac{1}{2}\log(ab) = \log\sqrt{ab}.$$

So the inequality reads

$$\log\left(\frac{a+b}{2}\right) \geq \log\sqrt{ab}, \tag{27.9}$$

with equality if and only if $a = b$. Since log is a strictly increang function, this inequality becomes

$$\frac{a+b}{2} \geq \sqrt{ab}, \tag{27.10}$$

with equality if and only if $a = b$. Now let us apply concavity to three values a, b, c; this leads to

$$\log\left(\frac{a+b+c}{3}\right) \geq \frac{1}{3}(\log a + \log b + \log c) = \log(abc)^{1/3},$$

which means

$$\frac{a+b+c}{3} \geq (abc)^{\frac{1}{3}}, \tag{27.11}$$

with equality if and only if a, b and c are equal.

The inequalities (27.9) and (27.12) area purely algebraic, having nothing to do with log, and, at first sight, seem to have no relationship to each other. (The first inequaity (27.9) can be proved directl just by simple algebra: $(\sqrt{a}-\sqrt{b})^2 \geq 0$ implies $a+b-2\sqrt{ab} \geq 0$). However, both express a common idea: the *arithmetic mean (AM) is greater or equal to the geometric mean.* (GM) The general version of this AM-GM inequality is:

$$\frac{x_1 + \cdots + x_N}{N} \geq (x_1 \ldots x_N)^{\frac{1}{N}}, \tag{27.12}$$

with equality if and only if all the x_i are equal. This follows from the strict concavity of log.

Not all convex functions of interest are differentiable everywhere. Here are a few convex functions that are not differentiable at all points:

$$|x|, \quad x_+ = \max\{x, 0\}, \quad (x-K)_+ = \max\{x-K, 0\},$$

for any constant $K \in \mathbb{R}$.

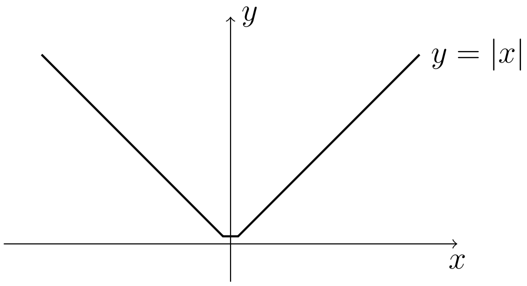


Figure 27.4: The convex function $|x|$

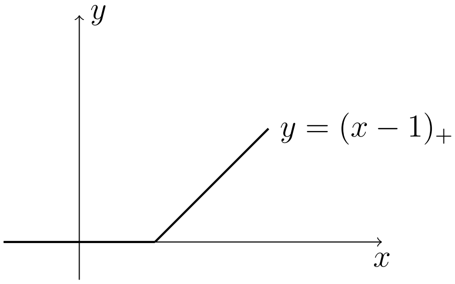


Figure 27.5: The convex function $(x-1)_+$

27.5 Convexity and derivatives

We have seen in section 27.2 that convexity of a function can be understood in terms of increasing secant slopes. From this we arrive at a condition for convexity for differentiable functions:

Proposition 27.5.1 *Suppose f is a function on an interval U on which f is differentiable.*

Then f is convex on U if and only if f' is an increasing function in the sense that $f'(s) \leq f'(t)$ for all $s, t \in U$ with $s \leq t$.

The function f is concave if and only if f' is an decreasing function in the sense that $f'(s) \geq f'(t)$ for all $s, t \in U$ with $s \leq t$.

Proof. Suppose f is convex on U, and $s, t \in U$ with $s < t$. Then for any $x, w \in [s, t]$ with

$$s < x < w < t$$

we have the increasing slope condition

$$\frac{f(x)-f(s)}{x-s} \leq \frac{f(t)-f(w)}{t-w}$$

Now taking the limits $x \to s$ and $w \to t$ we have

$$f'(s) \leq f'(t).$$

Conversely, suppose the function f on U has increasing slope f'. Consider $a, p, b \in U$ with

$$a < p < b.$$

Let A be the point $\bigl(a, f(a)\bigr)$, and B the point $(b, f(b))$ and P the point $(p, f(p))$. Then

$$\text{slope of } AP = \frac{f(p)-f(a)}{p-a} = f'(s) \quad \text{for some } s \in (a, p),$$

by the mean value theorem. We also have:

$$\text{slope of } PB = \frac{f(b)-f(p)}{b-p} = f'(t) \quad \text{for some } t \in (p, b),$$

Since $s \leq t$ (because p lies between them), we know that

$$f'(s) \leq f'(t).$$

Hence

$$\text{slope of } AP \leq \text{slope of } PB.$$

Since this holds for all points $a, p, b \in U$, with $a < p < b$, the function f is convex on U.

The argument for concavity is exactly similar. $\boxed{\text{QED}}$

Observe in the proof that if f' is assumed to be *strictly increasing*, that is

$$f(s) < f(t) \text{ whenever } s < t \text{ and } s, t \in U$$

then f is *strictly convex.*

Similarly, *if f' is strictly decreasing then f is strictly concave.*

When working with functions that are twice differentiable there is an easier condition for convexity:

Proposition 27.5.2 *Suppose f is a function on an interval U on which f is differentiable, and suppose also that f' is differentiable on U.*

Then

(i) *f is convex if and only if f'' is ≥ 0 on U;*

(ii) *f is concave if and only if f'' is ≤ 0 on U.*

For strict convexity/concavity we have:

(iii) *f is strictly convex if f'' is > 0 on U;*

(iv) *f is strictly concave if f'' is ≤ 0 on U.*

Note that in (iii) and (iv) we don't have the 'only if' parts. For example, x^4 is strictly convex but its second derivative is $12x^2$ which is 0 when $x = 0$. Proof. If $f'' \geq 0$ on U then f' is an increasing function on U (by Proposition 23.1.3) and so by Proposition 27.5.1 we conclude that f is convex. Conversely, if f is convex then by Proposition 27.5.1 the derivative f' is an increasing function on U and so, by Proposition 23.1.1, its derivative f'' is ≥ 0 on U. The results (iii) and (iv) follow similarly. $\boxed{\text{QED}}$

27.6 Supporting Lines

Consider a convex function Φ on an open interval $U \subset \mathbb{R}$. We have seen in (27.1) that for any point $p \in U$ the slopes of secant segments of the graph of f to the right of p exceed the slopes to the left of p¿ More formally:

$$\frac{\Phi(a) - \Phi(p)}{a - p} \leq \frac{\Phi(b) - \Phi(p)}{b - p} \quad \text{for all } a, b \in U \text{ with } a < p < b. \tag{27.13}$$

If we now squeeze in a value $m \in \mathbb{R}$ between these left secant segment slopes and right secant segment slopes (we will examine this more carefully below) then we have

$$\text{slope of any secant segment to the left of } p \leq m \leq \text{slope of any secant segment to the right of } p.$$

Consider now the line L through $(p, \Phi(p))$ whose slope is m. To the right of p it lies below all the secant segments for the graph of f and so *the line L lies below the graph of Φ to the right of p*. But to the left of p the slope of L is greater than the secant slopes and so again *the line L lies below the graph of Φ to the left of p*. Thus we have:

Proposition 27.6.1 *Suppose* Φ *is a convex function on an open interval* U*, and let* p *be any point in* U*. Then there is a value* $m \in \mathbb{R}$ *such that the line through* $\big(p, \Phi(p)\big)$ *with slope* m *lies below the graph of* Φ*; more precisely,*

$$L(x) \leq \Phi(x) \qquad \textit{for all } x \in U,$$

where $y = L(x)$ *is the equation of the line through* $(p, \Phi(p))$ *having slope* m*.*

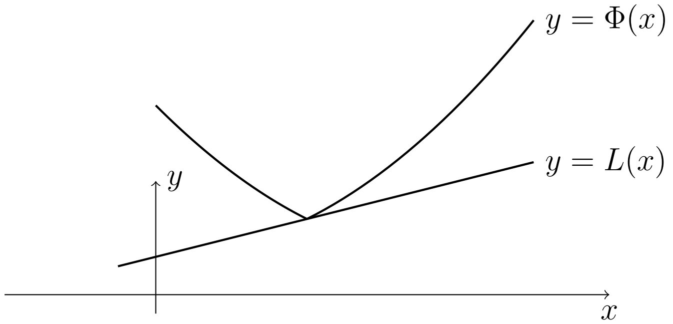


Figure 27.6: Supporting line for a convex function

A line such such as L is called a *supporting line* for Φ at p.

If $\Phi'(p)$ exists at p then there is only one choice for the line L: it is the line through $(p, \Phi(p))$ with slope $m = \Phi'(p)$, that is, it is the tangent line to the graph of Φ at $(p, \Phi(p))$. Notice that our conclusion that m is a real number forces $\Phi'(p)$, if it exists, to be finite.

Proof. All we have to do is prove that a real number m exists satisfying

$$\frac{\Phi(a) - \Phi(p)}{a - p} \leq m \leq \frac{\Phi(b) - \Phi(p)}{b - p} \quad \text{for all } a, b \in U \text{ with } a < p < b. \tag{27.14}$$

Since (27.13) holds for *every* $a \in U$ to the left of p we see that the 'right slope' $\frac{\Phi(b)-\Phi(p)}{b-p}$ is an upper bound for all the 'left slopes', and so

$$\sup_{a \in U, a < p} \frac{\Phi(a) - \Phi(p)}{a - p} \leq \frac{\Phi(b) - \Phi(p)}{b - p},$$

because sup is the *least upper bound*. Let us denote the sup by $\Phi'_{-}(p)$:

$$m_{-} \overset{\text{def}}{=} \sup_{a \in U, a < p} \frac{\Phi(a) - \Phi(p)}{b - p}.$$

Thus,

$$m_- \leq \frac{\Phi(b)-\Phi(p)}{b-p} \quad \text{for all } b \in U \text{ with } p < b.$$

Thus m_- is a lower bound for all the right slopes, and so

$$m_- \leq m_+ \tag{27.15}$$

where

$$m_+ \stackrel{\text{def}}{=} \sup_{b \in U, b>p} \frac{\Phi(b)-\Phi(p)}{b-p}.$$

Both the left and right derivatives $m_\pm$ are finite as they exceed all the secant segment slopes to the left of p and are $\leq$ all the secant segment slopes to the right of p. Now we can simply take m to lie between these values:

$$m_-leqm \leq m_+$$

$\boxed{\text{QED}}$ The secant slope inequalities for a convex function have the following remarkable consequence:

Proposition 27.6.2 *Any convex function on an open interval U is continuous on U.*

Proposition 27.6.1 leads to a way of understanding convex functions by means of the supporting lines:

Proposition 27.6.3 *If Φ is a convex function on an open interval U then Φ is the supremum of all the 'line-functions' that lie below it:*

$$\Phi(p) = \sup_{L \leq \Phi} L(p) \qquad \textit{for all } p \in U, \tag{27.16}$$

where L denotes any function of the form $L(x) = Mx + k$, with constant $M, k \in \mathbb{R}$, for all $x \in U$.

Proof. We have already seen in Proposition 27.6.1 that the graph of Φ has a supporting line at every point. Thus for any $p \in U$ there is a function L, whose graph is a line, for which $L(x) \leq \Phi(x)$ for all $x \in U$ (this means the graph of L lies below the graph of Φ) and $L(p) = \Phi(p)$. This proves (27.16), by showing that in fact there is actually an L for which both $L \leq \Phi$ on U and $L(p) = \Phi(p)$. $\boxed{\text{QED}}$

27.7 Convex combinations

If $a, b \in \mathbb{R}$ are points with $a < b$, then any point $p \in [a, b]$ can be reached by starting at a and moving towars b, covering a fraction

$$\frac{p-a}{b-a}$$

of the interval $[a, b]$. Thus,

$$p = a + \left[\frac{p-a}{b-a}\right](b-a),$$

a formula you can readily check directly with algebra. Let μ denote the fraction

$$\mu = \frac{p-a}{b-a}.$$

Since $p \in [a, b]$ it is clear that $\mu \geq 0$ and the most it can be is $(b-a)/(b-a) = 1$:

$$\mu \in [0, 1].$$

We can write p as

$$p = a + \mu(b-a) = a + \mu(b) - \mu a = a - \mu a\ +\ \mu b = (1-\mu)a + \mu b.$$

Writing λ for $1 - \mu$ we thus see that every point $p \in [a, b]$ can be expressed as a *convex combination* of a and b:

$$p = \lambda a + \mu b, \tag{27.17}$$

where λ and μ are *weights*, in the sense that they are non-negative and add up to 1 (which forces λ and μ to be ≤ 1):

$$\lambda, \mu \in [0, 1], \qquad \lambda + \mu = 1.$$

Conversely, any convex combination (27.17) is at most b (the weight on a, which is $< b$, would draw the value of p down below b) and at least a:

$$a \leq \text{any convex combination of } a \text{ and } b \leq b.$$

The point half way between a and b is

$$\frac{1}{2}a + \frac{1}{2}b = \frac{a+b}{2},$$

whereas the point 2/5-th of the way from a to b is:

$$\frac{3}{5}a + \frac{2}{5}b = \frac{23a + 2b}{5}.$$

A convex combination of points $p_1, \ldots, p_N$ is a point which can be expressed as

$$w_1 p_1 + \cdots + w_N p_N,$$

where the]em weights $w_1, \ldots, w_N$ all lie in $[0, 1]$ and sum to 1:

$$w_1 + \cdots + w_N = 1.$$

Note that if $p_1 = \cdots = p_N$, all points coinciding, then their convex combination is just that one point, since then

$$w_1 q + \cdots + w_N q = (w_1 + \cdots + w_N)q = 1 \cdot q = q,$$

where q is the common value of the p_i.

For a given set of points $p_1, \ldots, p_N$, the largest (the most to the right) that a convex combination could be is $\max_j p_j$, and for this we would have to distribute the entire weight 1 on those j for which p_j is largest, and take all the other weights to be 0; for example, for a, b, c, with$d = c > b > a$, the weighed average

$$w_1 a + w_2 b + w_3 c + w_4 d = w_1 a + w_2 b + (w_3 + w_4)d$$

is largest if $w_3 + w_4 = 1$ and $w_1 = w_2 = 0$.

Similarly, the least value a convex combintation of $p_1, \ldots, p_N$ could have is $\min_j p_j$, and this is obtained if and only if a weight of 0 is given to non-minimum values of p_j.

A multiple convex combination can be built out of convex combinations of pairs. For example,

$$\frac{2}{10}a + \frac{3}{10}b + \frac{5}{10}c = \frac{2}{10}a + \frac{8}{10}\left[\frac{\frac{3}{10}b + \frac{5}{10}c}{\frac{8}{10}}\right],$$

where notice that the quantity inside $[\cdots]$ is indeed also a convex combination; we thus have, on the right, a convex combination of convex combinations. More generally, for $N \in \{2, 3, \ldots\}$, any points $p_1, \ldots, p_N \in \mathbb{R}$, and any weights $w_1, \ldots, w_N \in [0, 1]$ (summing to 1) we have

$$w_1 p_1 + \cdots + w_N p_N = w_1 p_1 + (1 - w_1)\left[\frac{w_1 p_1 + \cdots + w_N p_N}{1 - w_1}\right] \qquad (27.18)$$

and the quantity inside $[\cdots]$ is a convex combination of the $N-1$ points $p_1, \ldots, p_{N-1}$.

The following result expresses the simultaneous convexity and concavity of functions whose graphs are straight lines:

Proposition 27.7.1 *If L is a function whose graph is a straight line, that is if*

$$L(x) = Mx + k \qquad \text{for all } x \in \mathbb{R},$$

for some constants $M, k \in \mathbb{R}$, then

$$L(w_1p_1 + \cdots + w_Np_N) = w_1L(p_1) + \cdots + w_NL(p_N) \tag{27.19}$$

for all $p_1, \ldots, p_N \in \mathbb{R}$ and $w_1, \ldots, w_N \in \mathbb{R}$ with $w_1 + \cdots + w_N = 1$.

Note that in (27.19) we have just a *linear combination*

$$w_1p_1 + \cdots + w_Np_N$$

and the coefficients w_i need not be in $[0,1]$ nor have to sum to 1.
Proof. It is more convenient to start with the right side of (27.19):

$$\begin{aligned} w_1L(p_1) + \cdots + w_NL(p_N) &= w_1(Mp_1 + k_1) \cdots + w_N(Mp_N + k) \\ &= w_1Mp_1 + w_1k + \cdots + w_NMp_N + w_Nk \\ &= Mw_1p_1 + \cdots + Mw_Np_N + (w_1 + \cdots + w_N)k \\ &= M(w_1p_1 + \cdots + w_Np_N) + k \\ &\qquad \text{(because } w_1 + \cdots + w_N = 1) \\ &= L(w_1p_1 + \cdots + w_Np_N). \end{aligned}$$

Thus L maps linear combinations to linear combinations. $\boxed{\text{QED}}$

This leads to the following convenient formulation of convexity of a function Φ:

Proposition 27.7.2 *A function Φ on an interval $U \subset \mathbb{R}$ is convex if and only if*

$$\Phi(\lambda a + \mu b) \leq \lambda\Phi(a) + \mu\Phi(b) \tag{27.20}$$

for all $a, b \in \mathbb{R}$ and all weights $\lambda, \mu \in [0,1]$ with $\lambda + \mu = 1$. The function Φ is strictly convex if and only if (27.20) holds for all a, b, λ, μ as above but with $\leq$ replaced by $<$ whenever the three points a, $\lambda a + \mu b$ and b are distinct (no two are equal to each other).

Proof. We can work with $a, b \in U$, with $a < b$ (if $a = b$ then (27.20) is an equality, both sides being $\Phi(a)$). Let A be the point $(a, \Phi(a))$ and B the point $(b, \Phi(b))$. The straight line joining A to be B has equation

$$y = L(x) = Mx + k$$

for some constants. Consider now any point $p \in [a, b]$; we can write this as

$$p = \lambda a + \mu b$$

for some $\lambda, \mu \in [0, 1]$ with $\lambda + \mu = 1$ (see (27.17)). The condition that the graph of Φ is below the graph of L is

$$\Phi(p) \leq L(p)$$

for all such p. Now

$$L(p) = L(\lambda a + \mu b) = \lambda L(a) + \mu L(b),$$

by Proposition 27.7.1. Since $y = L(x)$ passes through A and B, on the graph $y = \Phi(x)$, we have

$$L(a) = \Phi(a), \qquad \text{and} \qquad L(b) = \Phi(b). \tag{27.21}$$

Combining all these observations we have

$$\Phi(\lambda a + \mu b) \leq \lambda L(a) + \mu L(b) = \lambda\Phi(a) + \mu\Pi(b),$$

which establishes (27.20) as being equivalent to the convexity condition for Φ. For strict convexity, the point $\big(p, \Phi(p)\big)$ lies strictly below $(p, L(p))$, which means $\Phi(p) < L(p)$ when p is strictly between a and b. Translating from p to $\lambda a + \mu b$, and using again the equalities (27.21) we obtain the condition for strict convexity of Φ. $\boxed{\text{QED}}$

It is now easy to raise the inequality (27.20) to an inequality for convex combinations for multiples points. For example, for points $p_1, p_2, p_3 \in U$, we have

$$\begin{aligned}
\Phi(w_1p_1 + w_2p_2 + w_3p_3) &= \Phi\left(w_1p_1 + (1 - w_1)\left(\frac{w_2p_2 + w_3p_3}{1 - w_1}\right)\right) \\
&\leq w_1L(p_1) + (1 - w_1)L\left(\frac{w_2p_2 + w_3p_3}{1 - w_1}\right) \\
&= w_1L(p_1) + (1 - w_1)\left(\frac{w_2}{1 - w_1}L(p_2) + \frac{w_3}{1 - w_1}L(p_3)\right) \\
&= w_1L(p_1) + w_2L(p_2) + w_3L(p_3).
\end{aligned}$$

This procedure (the method of induction) leads to the conclusion that

$$\Phi(w_1p_1 + \cdots + w_Np_N) \leq w_1\Phi(p_1) + \cdots + w_N\Phi(p_N) \tag{27.22}$$

for every convex function Φ on any interval U, any $N \in \{1, 2, \ldots\}$, any points $p_1, \ldots, p_N \in U$, and all weights $w_1, \ldots, w_N \in [0, 1]$ adding up to 1.

Exercises on Maxima/Minima, Mean Value Theorem, Convexity

1. Find the maximum value of $x^{2/x}$ for $x \in (0, \infty)$. Explain your reasoning fully and present all calculations clearly.

2. Find the distance of the point $(1, 2)$ from the line whose equation is

$$3x + 4y - 5 = 0.$$

3. Suppose f is a twice differentiable function on $[1, 5]$, with $f(1) = f(3) = f(5)$. Show that there is a point $p \in (1, 5)$ where $f''(p)$ is 0.

4. Explain briefly why

$$\log 101 - \log 100 < .01.$$

5. Prove the inequality

$$\frac{1}{\left(\frac{a+b}{2}\right)^2} \leq \frac{1}{2}\frac{1}{a^2} + \frac{1}{2}\frac{1}{b^2}$$

for any $a, b > 0$.

Chapter 28

L'Hospital's Rule

L'Hospital's rule makes it possible to compute weird limits such as

$$\lim_{x\to\infty} x^{1/x},$$

and is worth studying just for that reason. It has been a staple topic in any introduction to calculus since l'Hospital's own book, reputed to be the first textbook on calculus.

Briefly, l'Hospital's rule says

$$\lim_{x\to p} \frac{f(x)}{g(x)} = \lim_{x\to p} \frac{f'(x)}{g'(x)} \tag{28.1}$$

if $f(x)$ and $g(x)$ both go to 0, or both go to $\pm\infty$, as $x \to p$, and if the limit on the right exists. Assuming that f and g both have domain a set S, here is a more detailed statement of the conditions:

- there is a neighborhood U of $p \in \mathbb{R}^*$ such that that the part of U in S, with p added in if necessary, that is the set $W = (S \cap U) \cup \{p\}$, is either U or a one-sided neighborhood of p of the form $(a, p]$ or $[p, a)$, for some $a \in \mathbb{R}$;
- $g(x) \neq 0$ and $g'(x) \neq 0$ for all $x \in W$, with $x \neq p$;
- $\lim_{x\to p} f(x)$ and $\lim_{x\to p} g(x)$ are either both 0 or are both in $\{-\infty, \infty\}$;
- the limit $\lim_{x\to p} \frac{f'(x)}{g'(x)}$ exists.

28.1 Examples

Avoiding silly examples such as

$$\lim_{x\to 1} \frac{x^2-1}{x-1}$$

(which is 2 as can be seen directly from $x^2 - 1 = (x-1)(x+1)$) let us work out the following simple but more interesting use of l'Hospital's rule:

$$\lim_{x\to 0} \frac{\sin x - x}{\frac{1}{3!}x^3}.$$

The first thing to observe is that both numerator and denominator go to 0 as $x \to 0$. So we could, poentially use l'Hospital's rule:

$$\lim_{x\to 0} \frac{\sin x - x}{\frac{1}{3!}x^3} = \lim_{x\to 0} \frac{\cos x - 1}{\frac{1}{3!}3x^2}, \tag{28.2}$$

provided the limit on the right exists. To deal with this limit observe again that both numerator and denominator go to 0 as $x \to 0$, and so we could again try l'Hospital:

$$\lim_{x\to 0} \frac{\cos x - 1}{\frac{1}{3!}3x^2} = \lim_{x\to 0} \frac{-\sin x}{\frac{1}{3!}3\cdot 2\,x} = -\lim_{x\to 0} \frac{\sin x}{x}, \tag{28.3}$$

and *we do know that this limit exists* and its value is -1. Thus l'Hospital's rule does imply that the equality (28.3) holds, and this shows that the right side of (28.2) exists, which then justifies the equality (28.2) by l'Hospital's rule. This somewhat convoluted logic is summarized simply in:

$$\begin{aligned}
\lim_{x\to 0} \frac{\sin x - x}{\frac{1}{3!}x^3} &\overset{\text{l'H.}}{=} \lim_{x\to 0} \frac{\cos x - 1}{\frac{1}{3!}3x^2} \quad \text{(if the right side exists)} \\
&\overset{\text{l'H.}}{=} \lim_{x\to 0} \frac{-\sin x}{\frac{1}{3!}3\cdot 2\,x} \quad \text{(if this exists)} \\
&= -\lim_{x\to 0} \frac{\sin x}{x} \\
&= -1 \quad \text{(whch justifies the 2nd and hence the 1st equality above).}
\end{aligned} \tag{28.4}$$

Thus, for x close to 0 we should be able to approximate the difference $\sin x - x$ by $-\frac{1}{3!}x^3$, and so

$$\sin x \simeq x - \frac{1}{3!}x^3$$

for x near 0.

Now we carry this a step beyond, finding an estimate for the difference

$$\sin x - \left[x - \frac{1}{3!}x^3\right].$$

By repeated use of l'Hospital's rule we have:

$$\begin{aligned}
\lim_{x\to 0} \frac{\sin x - \left[x - \frac{1}{3!}x^3\right]}{\frac{1}{5!}x^5} &\overset{\text{l'H.}}{=} \lim_{x\to 0} \frac{\cos x - \left[1 - \frac{1}{3!}3x^2\right]}{\frac{1}{5!}x^5} \quad \text{(if the right side exists)}\\
&= \lim_{x\to 0} \frac{\cos x - \left[1 - \frac{1}{2!}x^2\right]}{\frac{1}{4!}x^4} \quad \text{(by algebraic simplification)}\\
&\overset{\text{l'H.}}{=} \lim_{x\to 0} \frac{-\sin x - \left[-\frac{1}{2!}2x\right]}{\frac{1}{4!}4x^3} \quad \text{(if the right side exists)}\\
&= -\lim_{x\to 0} \frac{\sin x - x}{\frac{1}{3!}x^3} \quad \text{(by algebraic simplification)}\\
&= -(-1) \quad \text{(by the previous example)}\\
&= 1.
\end{aligned} \tag{28.5}$$

Each application of the l'Hospital rule above was a case where the limit of the ratio had the form 0/0.

We turn now to a different example:

$$\begin{aligned}
\lim_{x\to\infty} x^{\frac{1}{x}} &= \lim_{x\to\infty} \left[e^{\log x}\right]^{\frac{1}{x}}\\
&= \lim_{x\to\infty} e^{\frac{\log x}{x}}\\
&= e^{\lim_{x\to\infty} \frac{\log x}{x}} \qquad \text{(the exponent here is formally } \infty/\infty)\\
&\overset{\text{l'H.}}{=} e^{\lim_{x\to\infty} \frac{1/x}{1}}\\
&= e^0\\
&= 1.
\end{aligned} \tag{28.6}$$

28.2 Proving l'Hospital's rule

The key step in proving l'Hospital

$$\lim_{x\to p}\frac{f(x)}{g(x)}=\lim_{x\to p}\frac{f('x)}{g'(x)},$$

with $f(x)$ and $g(x)$ both $\to 0$ as $x \to p$, is the observation that

$$\frac{f(x)}{g(x)}=\frac{f('c)}{g'(c)},$$

for some c between x and p; when $x \to p$, the point c also $\to p$ and this shows that the above ratios approach the same limit. This is formalized in the following version of the mean value theorem:

Proposition 28.2.1 *Suppose F and G are continuous functions on a closed interval $[a,b]$, where $a,b \in \mathbb{R}^*$ and $a<b$, with values in $\mathbb{R}$. Suppose that F and G are differentiable on (a,b), with $G'(x) \neq 0$ for all $x \in (a,b)$. Then*

$$\frac{F(b)-F(a)}{G(b)-G(a)}=\frac{F'(c)}{G'(c)} \tag{28.7}$$

for some $c \in (a,b)$.

Since G' is never 0 on (a,b) it follows by Rolle's theorem that $G(b)-G(a) \neq 0$.
Proof. Consider the function H defined on $[a,b]$ by

$$H(x)=[G(b)-G(a)]\,[F(x)-F(a)]-[F(b)-F(a)]\,[G(x)-G(a)] \quad \text{for all } x \in [a,b]. \tag{28.8}$$

This is clearly continuous on $[a,b]$ and differentiable on (a,b) with derivative given by

$$H('x)=[G(b)-G(a)]\,F'(x)-[F(b)-F(a)]\,G'(x) \tag{28.9}$$

for all $x \in (a,b)$.

Observe also that

$$H(a)=H(b)=0.$$

Then by Rolle's theorem applied to H there is a point $c \in (a, b)$ where $H'(c)$ is 0; this means

$$[G(b) - G(a)]\,F'(c) - [F(b) - F(a)]\,G'(c) = 0. \tag{28.10}$$

Thus

$$[F(b) - F(a)]\,G'(c) = [G(b) - G(a)]\,F'(c),$$

which implies the result (28.7). $\boxed{\text{QED}}$

Now we can prove one form of l'Hospital's rule:

Proposition 28.2.2 *Suppose f and g are differentiable functions on an interval $U \subset \mathbb{R}^*$, with $g(x) \neq 0$ and $g'(x) \neq 0$ for all $x \in U$ in some neighborhood of a limit point $p \in \mathbb{R}^*$ of U, and suppose*

$$\lim_{x\to p} f(x) = 0 \qquad \textit{and} \qquad \lim_{x\to p} g(x) = 0.$$

Then

$$\lim_{x\to p} \frac{f(x)}{g(x)} = \lim_{x\to p} \frac{f'(x)}{g'(x)} \tag{28.11}$$

if the limit on the right in (28.11) exists.

Proof. Let W be a closed interval, one of whose endpoints is p and for which all points of W, except possibly for p, lie inside U. Define F and G on $U \cup \{p\}$ by requiring that $F(x) = f(x)$ and $G(x) = g(x)$ for all $x \in U$ with $x \neq p$, and setting

$$F(p) = 0 \qquad \text{and} \qquad G(p) = 0.$$

What this does is make f and g defined and continuous at p in case it wasn't to start with; more precisely, F and G are continuous on $U \cup \{p\}$. For any $x \in U$, in a neighborhood of p, with $x \neq p$ we have

$$\frac{f(x)}{g(x)} = \frac{F(x) - F(p)}{G(x) - G(p)},$$

because $F(p)$ and $G(p)$ are both 0 and $F(x) = f(x)$ and $G(x) = g(x)$. Then, for such x, we have by Proposition 28.2.1

$$\frac{f(x)}{g(x)} = \frac{F'(c_x)}{G'(c_x)} = \frac{f'(c_x)}{g'(c_x)},$$

for some c_x strictly between x and p. Letting $x \to p$ makes $c_x \to p$ and so

$$\lim_{x\to p}\frac{f(x)}{g(x)} = \lim_{x\to p}\frac{f'(c_x)}{g'(c_x)} = \lim_{w\to p}\frac{f'(w)}{g'(w)},$$

since we assume that the right side here exists. $\boxed{\text{QED}}$

We turn now to the case where f and g have infinite limits:

Proposition 28.2.3 *Suppose f and g are differentiable functions on an interval $U \subset \mathbb{R}^*$, with $g(x) \neq 0$ and $g'(x) \neq 0$ for all $x \in U$ in some neighborhood of a limit point $p \in \mathbb{R}^*$ of U, and suppose*

$$\lim_{x\to p} f(x) \in \{-\infty, \infty\} \qquad \textit{and} \qquad \lim_{x\to p} g(x) \in \{-\infty, \infty\}.$$

Then

$$\lim_{x\to p}\frac{f(x)}{g(x)} = \lim_{x\to p}\frac{f'(x)}{g'(x)} \tag{28.12}$$

if the limit on the right in (28.12) exists.

Proof. From

$$\frac{f(x)-f(a)}{g(x)-g(a)} = \frac{f(x)}{g(x)}\left[\frac{1-\frac{f(a)}{f(x)}}{1-\frac{g(a)}{g(x)}}\right] \tag{28.13}$$

we see that since both $f(x)$ and $g(x)$ approach $\pm\infty$ as $x \to p$, the limiting behavior of $f(x)/g(x)$ and $\frac{f(x)-f(a)}{g(x)-g(a)}$ is the same. This is the motivation for the strategy we use.

Let

$$L = \lim_{x\to p}\frac{f'(x)}{g'(x)},$$

and W any neighborhood of L. Choose a smaller neighborhood W_1 of L such that

$$cW_1 \subset W \qquad \text{for all } c \in (1/2, 2). \tag{28.14}$$

There is an interval V that is a neighborhood of p such that

$$\frac{f'(x)}{g'(x)} \in W_1 \tag{28.15}$$

for all $x \in V \cap U$ and $x \neq p$. Pick any two distinct points $a, x \in V \cap U$, neither equal to p; then

$$\frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f'(c)}{g'(c)}$$

for some c between a and x, and hence in V (note that $g(x) \neq g(a)$ because by Rolle's theorem if $g(x) = g(a)$ then g' would be 0 at a point between a and x, contrary to the given assumption that g' is never 0 on U). Hence by (28.15) we have

$$\frac{f(x) - f(a)}{g(x) - g(a)} \in W_1,$$

for distinct $a, x \in V \cap U$, neither equal to p. Since

$$\lim_{x \to p} \left[\frac{1 - \frac{g(a)}{g(x)}}{1 - \frac{f(a)}{f(x)}} \right] = 1$$

there is a neighborhood U_0 of p such that

$$\frac{1 - \frac{g(a)}{g(x)}}{1 - \frac{f(a)}{f(x)}} \in (1/2, 2) \qquad \text{for all } x \in U_0 \cap U$$

From (28.13) we have

$$\frac{f(x)}{g(x)} = \left[\frac{f(x) - f(a)}{g(x) - g(a)} \right] \left[\frac{1 - \frac{g(a)}{g(x)}}{1 - \frac{f(a)}{f(x)}} \right]$$

and so, on using (28.14), we conclude that

$$\frac{f(x)}{g(x)} \in W$$

for all $x \in U_0 \cap U \cap V$. Since this is true for any neighborhood W of L, we conclude that $f(x)/g(x) \to L$ as $x \to p$. $\boxed{\text{QED}}$

Exercises on l'Hosptal's rule

1. Work out the limit
$$\lim_{x\to 0}\frac{\cos x - \left[1-\frac{1}{2!}x^2\right]}{\frac{1}{4!}x^4}$$
clearly justifying each step.

2. Suppose g' is continuous, $g(2)=0$ and $g'(2)=1$. Work out the limit
$$\lim_{w\to 0}\frac{g(2+w)+g(2+3w)}{w}$$
clearly justifying each step.

3. Find
$$\lim_{y\to\infty} y^{\frac{1}{y}}$$

Chapter 29

Integration

The development of calculus has two original themes: (i) the notion of tangent to a curve, (ii) computing areas of curved regions. The former leads to differential calculus, and the latter to integral calculus, to which we turn now.

29.1 From areas to integrals

The classical idea of area A of a region S enclosed by a curve is that A should be $\leq$ the sum of areas of any finite collection of squares that cover the region S and A should be $\geq$ the sum of areas of any finite collection of squares inside S. This simple and perfectly natural idea fails to produce a completely satisfactory and usable measure of area when the region S is 'unintuitive' (for example S consists of all points $(x, y) \in \mathbb{R}^2$ with irrational coordinates), but it is meaningful, intuitive and computable for regions bounded by well behaved curves.

We are mainly interested in the area of the region lying below a graph

$$y = f(x),$$

and above the x-axis, for $x \in [a, b]$. See Figure 29.1. For this discussion we assume $f \geq 0$. Suppose A is the area of this region, a notion we will pin down as the discussion progresses.

An overestimate of A will surely be obtained by the sum of areas of 'upper rectangles' obtained by slicing the region into vertical pieces. More precisely partition $[a, b]$ into N intervals marked off by

$$a = t_0 < t_1 < \ldots < t_N = b.$$

By the k-th 'upper rectangle' we mean the rectangle whose base runs along the x-axis from $x = t_{k-1}$ to $x = t_k$ and whose height is

$$M_k = \sup_{x \in [t_{k-1}, t_k]} f(x). \tag{29.1}$$

The area of this upper rectangle is

$$M_k(t_k - t_{k-1}).$$

The width $t_k - t_{k-1}$ is often denoted by Δt_k:

$$\Delta t_k = t_k - t_{k-1}. \tag{29.2}$$

Thus the area of the k-th upper rectangle is

$$M_k \Delta t_k.$$

The sum of all the upper rectangles is

$$\sum_{k=1}^{N} M_k \Delta t_k = M_1 \Delta t_1 + \cdots + M_N \Delta t_N.$$

Thus the *upper rectangles* provide an *overestimate* of the area A:

$$A \leq \sum_{k=1}^{N} M_k \Delta t_k.$$

Similarly, working with *lower rectangles* we have an *understimate* of the area:

$$\sum_{k=1}^{N} m_k \Delta t_k \leq A,$$

where

$$m_k = \inf_{x \in [t_{k-1}, t_k]} f(x). \tag{29.3}$$

Thus the actual area A lies between the overestimates given by the *upper sums* and underestimates given by the *lower sums*:

$$\sum_{k=1}^{N} m_k \Delta t_k \leq A \leq \sum_{k=1}^{N} M_k \Delta t_k. \tag{29.4}$$

Surely, the area A is *the unique value that lies between all the upper sums and all the lower sums.* We can take this as definition of area under a graph.

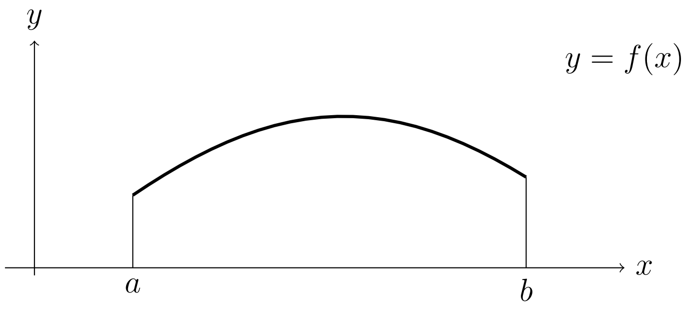
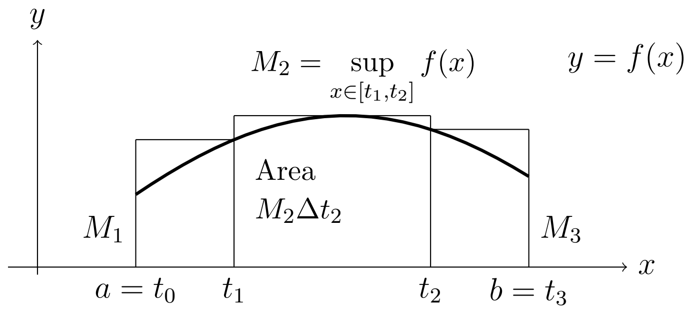


Figure 29.1: Area below $y = f(x)$



Area under graph is $\leq \underbrace{M_1\Delta t_1 + M_2\Delta_2 + M_3\Delta t_3}_{\text{sum of upper rectangle areas}}$

Figure 29.2: Area below $y = f(x)$ overestimated by upper recangles

29.2 The Riemann integral

The ideas of the preceding section lead to the crucial notion of the integral of a function. Let us first extract some helpful terminology from our earlier discussions.

A *partition* P of an interval $[a, b]$, with $a, b \in \mathbb{R}$ and $a \leq b$, is a finite subset of $[a, b]$ containing both the end points a and b. Typically we denote a partition by

$$P = \{x_0, x_1, \ldots, x_N\},$$

where

$$a = x_0 < x_1 < \ldots < x_N = b.$$

The width of the k-th interval is denoted

$$\Delta x_k = x_k - x_{k-1}. \tag{29.5}$$

For a function

$$f : [a, b] \to \mathbb{R},$$

and the partition P, the *upper sum* is

$$U(f, P) = \sum_{k=1}^{N} M_k(f)\Delta x_k, \tag{29.6}$$

and the *lower sum* is

$$L(f, P) = \sum_{k=1}^{N} m_k(f)\Delta x_k, \tag{29.7}$$

where

$$\begin{aligned} M_k(f) &= \sup_{x\in[t_{k-1},t_k]} f(x) \\ m_k(f) &= \inf_{x\in[t_{k-1},t_k]} f(x). \end{aligned} \tag{29.8}$$

In the degenerate case where $b = a$, the only partition of $[a, a]$ is just the one-point set $\{a\}$, and the upper and lower sums are taken to be 0.

If there is a unique value A for which

$$L(f, P) \leq A \leq U(f, P) \tag{29.9}$$

for every partition P of $[a, b]$, then A is called the *Riemann integral* of f, and denoted

$$\int_a^b f.$$

We will refer to this simply as *the integral of f from a to b* or *over* $[a, b]$.

We say that f is *integrable* if $\int_a^b f$ exists and is finite (in $\mathbb{R}$).

The definition of the integral here is in the same spirit that of the concept of limit back in (6.1) and the concept of tangent line in (13.1).

From (29.9) we see that an approximation to $\int_a^b f($ is given by

$$\int_a^b f(x)\,dx \simeq \sum_{k=1}^{N} f(x_k^*)\Delta x_k, \tag{29.10}$$

where x_k^* is any point in $[x_{k-1}, x_k]$, for each $k \in \{1, \ldots, N\}$. The sum on the right in (29.10) is called a *Riemann sum* for f with respect to the partition P. The relation (29.10) suggests the historical origin of the notation $\int$ for integration.

Note that if $\sup_{x\in[a,b]} f(x)$ is ∞ then at least one $M_k(f)$ is ∞, for any partition P, and so the upper sums are all ∞, and in this case the integral $\int_a^b f$, if it exists, must also be ∞. Similarly, if $\inf_{x\in[a,b]} f(x)$ is $-\infty$ then $\int_a^b f$, if it exists, is $-\infty$.

Thus, *if f is integrable then it is bounded,* in the sense that both its supremum and its infinmum are finite.

If f is constant, with value $C \in \mathbb{R}$ then, working with any partition P as above, we have

$$M_k(f) = C \qquad \text{and} \qquad m_k(f) = C,$$

for all $k \in \{1, \ldots, N\}$, and so

$$U(f, P) = C\Delta x_1 + \cdots + C\Delta x_N = C(\Delta x_1 + \cdots + \Delta x_N) = C(b - a),$$

and

$$L(f, P) = C\Delta x_1 + \cdots + C\Delta x_N = C(\Delta x_1 + \cdots + \Delta x_N) = C(b - a).$$

Hence

$$\int_a^b C = C(b - a). \tag{29.11}$$

29.3 Refining partitions

Consider a function f on an interval $[a, b]$, and a partition $P = \{x_0, \ldots, x_N\}$ of $[a, b]$, with

$$a = x_0 < \ldots < x_N = b.$$

Let us see what effect there is on the upper and lower sums when points are added to P to make it a finer partition of $[a, b]$. Let us start by adding one point $s \in (x_{j-1}, x_j)$ to the j-th interval.

In the sum

$$U(f, P) = \sum_{k=1}^{N} M_k(f)\Delta x_k$$

all terms remain the sum except for the j-th term: for this

$$M_j(f)\Delta x_j$$

is replaced by

$$A(s - x_{j-1}) + B(x_j - s)$$

where A is the sup of f over $[x_{j-1}, s]$ and B is the sup of f over $[s, x_j]$:

$$A = \sup_{x\in[x_{j-1},s]} f(x) \qquad \text{and} \qquad B = \sup_{x\in[s,x_j]} f(x).$$

Clearly these are both $\leq M_j(f)$:

$$A, B \leq M_j(f),$$

and in fact at least one of them is equal to $M_j(f)$. Hence

$$A(s - x_{j-1}) + B(x_j - s) \leq M_j(f)(s - x_{j-1}) + M_j(f)(x_j - s),$$

and observe that the right side here adds up to $M_j(f)\Delta x_j$; thus:

$$A(s - x_{j-1}) + B(x_j - s) \leq M_j(f)\Delta x_j.$$

Looking back at the upper sum $U(f, P)$ we conclude that

$$U(f, P_1) \leq U(f, P),$$

where P_1 is the partition obtained by adding the point s to P:

$$P' = P \cup \{s\}.$$

Thus, *adding a point to a partition* (that is, splitting one of the intervals into two) *lowers the upper sum.*

A similar argument shows that

$$L(f, P_1) \geq L(f, P);$$

adding a point to a partition raises the lower sum.

Adding points one by one enlarges a given partition P to any given larger partition P', and at each stage in this process the upper sum is lowered and the lower sum is raised:

Proposition 29.3.1 *Let $f : [a, b] \to \mathbb{R}$ be a function, where $a, b \in \mathbb{R}$ and $a \leq b$, and P and P' any partitions of $[a, b]$ with $P \subset P'$; then*

$$\begin{aligned} L(f, P) &\leq L(f, P') \\ U(f, P') &\leq U(f, P). \end{aligned} \tag{29.12}$$

This implies the following natural but strong observation:

Proposition 29.3.2 *Let $f : [a, b] \to \mathbb{R}$ be a function, where $a, b \in \mathbb{R}$ and $a \leq b$, and P and Q any partitions of $[a, b]$; then*

$$L(f, P) \leq U(f, Q). \tag{29.13}$$

Thus, every upper sum of f is $\geq$ every lower sum of f.

We have seen something similar in our study of limits back in (6.14). Proof. Let

$$P' = P \cup Q.$$

Then P' contains both P and Q, and so by Proposition 29.3.1 we have

$$L(f, P) \leq L(f, P') \qquad \text{and} \qquad U(f, P') \leq U(f, Q).$$

Combining this with the fact that $L(f, P') \leq U(f, P')$ produces the inequality (29.13). $\boxed{\text{QED}}$

29.4 Estimating approximation error

Consider a function f on an interval $[a, b] \subset \mathbb{R}$, and let $P = \{x_0, \ldots, x_N\}$ be a partition of $[a, b]$ with

$$a = x_0 < \ldots < x_N = b.$$

We know that the integral of f, if it exists, lies between the upper sum $U(f, P)$ and the lower sum $L(f, P)$. So if $U(f, P)$ and $L(f, P)$ are close to each other then either of these sums would be a good approximation to the value of the integral. Let us find how far from each other the upper and lower sums are:

$$\begin{aligned} U(f, P) - L(f, P) &= \sum_{k=1}^{N} M_k(f)\Delta x_k - \sum_{k=1}^{N} m_K(f)\Delta x_k \\ &= \sum_{k=1}^{N} \left[M_k(f) - m_k(f)\right] \Delta x_k, \end{aligned} \tag{29.14}$$

where $M_k(f)$ is the sup of f over $[x_{k-1}, x_k]$ and $m_k(f)$ is the inf of f over $[x_{k-1}, x_k]$. Thus

$$M_k(f) - m_k(f) = \text{the fluctuation of } f \text{ over } [x_{k-1}, x_k]. \tag{29.15}$$

Thus if we can partition $[a, b]$ so finely that the fluctuation of f over each interval $[x_{k-1}, x_k]$ is $< .01$ then

$$\begin{aligned} U(f, P) - L(f, P) &< .01\Delta x_1 + \cdots + .01\Delta x_N \\ &= .01\,[\Delta x_1 + \cdots + \Delta x_N] = (0.01)(b - a). \end{aligned} \tag{29.16}$$

Thus we can shrink $U(f, P) - L(f, P)$ down by choosing the partition P such that the fluctuation of f over each interval $[x_{k-1}, x_k]$ is very small.

29.5 Continuous functions are integrable

The discussions of the preceding section lead to the following important result:

Theorem 29.5.1 *If $f : [a, b] \to \mathbb{R}$ is continuous, where $a, b \in \mathbb{R}$ with $a \leq b$, then f is integrable.*

The key to proving this result is the ability to partition $[a, b]$ in such a way that the fluctuation of f over the intervals are all very small:

Proposition 29.5.1 *If $f : [a, b] \to \mathbb{R}$ is continuous, where $a, b \in \mathbb{R}$ with $a \leq b$, then for any $\epsilon > 0$ there is a partition $P = \{x_0, \ldots, x_N\}$ of $[a, b]$, with $a = x_0 < \ldots < x_N = b$, such that the fluctuation of f over each interval $[x_{k-1}, x_k]$ is $< \epsilon$.*

This property is called *uniform continuity* of f.
Proof. We work with $a < b$, as the case $a = b$ is trivial. Take

$$x_0 = a.$$

Since f is continuous at a there is some $x_1 > a = x_0$ such that

$$f(x_0) - \frac{\epsilon}{4} < f(x) < f(x_0) + \frac{\epsilon}{4}$$

for all $x \in [x_0, x_1]$. We can clearly take $x_1 \leq b$, as here is no need to go beyond b. Thus

$$\sup_{x \in [x_0, x_1]} f(x) \leq f(x_0) + \frac{\epsilon}{4}$$

and

$$\inf_{x \in [x_0, x_1]} f(x) \geq f(x_0) - \frac{\epsilon}{4}.$$

These conditions imply that the fluctuation of f over $[x_0, x_1]$ is $\leq \epsilon/2$, which is, of course, $< \epsilon$.

Now we can start at x_1, if it isn't already b, and produce a point $x_2 > x_1$ for which

$$f(x_1) - \frac{\epsilon}{4} < f(x) < f(x_1) + \frac{\epsilon}{4}$$

for all $x \in [x_1, x_2]$. Again, we can take $x_2 \leq b$. It might seem that in this way we could produce the desired partition P. But there could be a problem: the process might continue infinitely without reaching b. Fortunately, we can show that this will not happen.

Suppose s is the supremum of all $t \in [a, b]$ such that $[a, t]$ has a partition $P_0 = \{x_0, \ldots, x_K\}$ for which the fluctuations of f over every interval of the partition is $< \epsilon$. Note that $s > a$. By continuity of f at s there is an interval (p, q), centered at s, such that the fluctuation of f over $(p, q) \cap [a, b]$ is $< \epsilon$. Pick any point $t \in (p, s)$, with $t > a$; then since $t < s$ the definition of s implies that there is a partition

$$P_0 = \{x_0, \ldots, x_K\}$$

of $[a, t]$ such that the fluctuation of f over each interval $[x_{j-1}, x_j]$ is $< \epsilon$. Now pick any point $r \in [s, q) \cap [a, b]$ and set

$$x_{K+1} = r.$$

Since the fluctuation of f over (p, q) is $< \epsilon$, the fluctuation of f over the subinterval $[t, r]$ is $< \epsilon$. Thus we have produced a point r, which is $\geq s$, such that there is a partition $P = \{x_0, \ldots, x_{K+1}\}$ of $[a, r]$ for which the fluctuations of f are all $< \epsilon$. To avoid a contradiction with the definition of s, we must have $s = b$ (for otherwise, if $s < b$, we could have chosen r to be $> s$) and the partition P has the desired fluctuation property. $\boxed{\text{QED}}$

Now we can prove Theorem 29.5.1.

Proof of Theorem 29.5.1. All we need to do is show that for any $\epsilon > 0$ there is partition P of $[a,b]$ for which $U(f,P) - L(f,P)$ is $< \epsilon$; this will imply that there is a unique value that lies between all the upper sums and all the lower sums. Let $\epsilon > 0$ and set, for convenience

$$\epsilon_1 = \frac{\epsilon}{b-a}.$$

By Proposition 29.5.1 there is a partition

$$P = \{x_0, \ldots, x_N\},$$

of $[a,b]$ such that

$$a = x_0 < \ldots < x_N = b$$

and

$$M_k(f) - m_k(f) < \epsilon_1$$

for all $k \in \{1, \ldots, N\}$. Then, by the argument used before in (29.16), we have

$$U(f,P) - L(f,P) < \epsilon_1(b-a).$$

Our choice of ϵ_1 then means that $U(f,P) - L(f,P)$ is $< \epsilon$. $\boxed{\text{QED}}$

29.6 A function for which the integral does not exist

Consider the indicator function $1_{\mathbb{Q}}$ of the rationals:

$$1_{\mathbb{Q}}(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q}; \\ 0 & \text{if } x \notin \mathbb{Q} \end{cases} \tag{29.17}$$

We focus on the restriction of $1_{\mathbb{Q}}$ over any interval $[a,b] \subset \mathbb{R}$, with $a < b$. Let $P = \{x_0, \ldots, x_N\}$ be any partition of $[a,b]$, with

$$a = x_0 < x_1 < \ldots < x_N = b.$$

Observe that $1_{\mathbb{Q}}$ attains both the value 1 and the value 0 on each interval $[x_{k-1}, x_k]$ (because every such interval contains both rational and irrational points). Hence

$$M_k(1_{\mathbb{Q}}) = \sup_{x \in [x_{k-1}, x_k]} 1_{\mathbb{Q}}(x) = 1 \quad \text{and} \quad m_k(1_{\mathbb{Q}}) = \inf_{x \in [x_{k-1}, x_k]} 1_{\mathbb{Q}}(x) = 0.$$

This makes the upper sum large:

$$U(1_{\mathbb{Q}}, P) = 1 \cdot \Delta x_1 + \ldots + 1 \cdot \Delta x_N = b - a,$$

and the lower sum small:

$$L(1_{\mathbb{Q}}, P) = 0 \cdot \Delta x_1 + \ldots + 0 \cdot \Delta x_N = 0.$$

There certainly are many real numbers lying between 0 and $b - a$, and so there is no unique such choice. Hence,

$$\int_a^b 1_{\mathbb{Q}} \quad \text{does not exist.}$$

29.7 Basic properties of the integral

Integration of a larger function produces a larger number:

Proposition 29.7.1 *If f and g are functions on an interval $[a, b]$, where $a, b \in \mathbb{R}$ and $a \leq b$, for which the integrals $\int_a^b f$ and $\int_a^b g$ exist, and if $f \leq g$ then*

$$\int_a^b f \leq \int_a^b g.$$

Proof. Suppose $\int_a^b f > \int_a^b g$. Since $\int_a^b g$ is the *unique* value lying between $L(g, P)$ and $U(g, P)$ for all partitions P of $[a, b]$, there must be a partition P of $[a, b]$ such that

$$\int_a^b f > U(g, P).$$

Again, since $\int_a^b f$ is the unique value lying between all upper and lower sums for f there is a partition Q of $[a, b]$ for which

$$L(f, Q) > U(g, P). \tag{29.18}$$

Let

$$P' = P \cup Q,$$

which is a partition of $[a, b]$. Since $Q \subset P'$ and $f \leq g$ we have

$$L(f, Q) \leq L(g, Q) \leq L(g, P').$$

Since $P \subset P'$ we also have

$$U(g, P') \leq U(g, P).$$

Combining these inequalities produces

$$L(f, Q) \leq L(g, P),$$

contradicting (29.18). $\boxed{\text{QED}}$

Next we have linearity of the integral:

Proposition 29.7.2 *Suppose f and g are functions on an interval $[a, b]$, where $a, b \in \mathbb{R}$ and $a \leq b$, for which the integrals $\int_a^b f$ and $\int_a^b g$ exist. Assume also that the sum*

$$\int_a^b f + \int_a^b g$$

is defined (thus, not $\infty + (-\infty)$ or $(-\infty) + \infty$). Then the integral $\int_a^b f$ exists and

$$\int_a^b (f + g) = \int_a^b f + \int_a^b g. \tag{29.19}$$

Moreover, for any $k \in \mathbb{R}$ the integral $\int_a^b kf$ of the function kf also exists and

$$\int_a^b kf = k \int_a^b f. \tag{29.20}$$

If a function is integrable over an interval then it is integrable over any subinterval:

Proposition 29.7.3 *Suppose f is integrable over $[a, b]$, where $a, b \in \mathbb{R}$ and $a \leq b$. The f is integrable over $[c, d]$ for any $c, d \in [a, b]$ with $c \leq d$. Moreover,*

$$\int_a^b f = \int_a^c f + \int_c^b f. \tag{29.21}$$

Chapter 30

The Fundamental Theorem of Calculus

The fundamental theorem of calculus connects differential calculus and integral calculus, and makes it possible to compute integrals by running the derivative process in reverese.

30.1 Fundamental theorem of calculus

Here is one form of the fundamental theorem:

Theorem 30.1.1 *Suppose f is an integrable function on $[a, b]$, where $a, b \in \mathbb{R}$, with $a < b$. Then the function F defined on $[a, b]$ by*

$$F(x) = \int_a^x f \qquad \textit{for all } x \in [a, b]$$

is differentiable at $p \in [a, b]$ if f is continuous at p, and $F'(p)$ is $f(p)$. Thus, if f is continuous on $[a, b]$ then

$$\frac{d \int_a^x f}{dx} = f(x) \tag{30.1}$$

for all $x \in [a, b]$.

Notice that (30.1) implies that *any continuous function f on an interval is the derivative of some function F on that interval.*

Here is another version of the result:

Theorem 30.1.2 *Suppose g is a differentiable function on $[a,b]$, where $a, b \in \mathbb{R}$, with $a < b$, for which g' is continuous. Then*

$$\int_a^b g' = g\Big|_a^b. \tag{30.2}$$

where $g\Big|_a^b$ means $g(b) - g(a)$:

$$g\Big|_a^b = g(b) - g(a).$$

30.2 Differentials and integrals

Consider a function f, differentiable at a point $p \in \mathbb{R}$. The tangent line to the graph $y = f(x)$ through the point $\big(p, f(p)\big)$ has slope $f'(p)$.

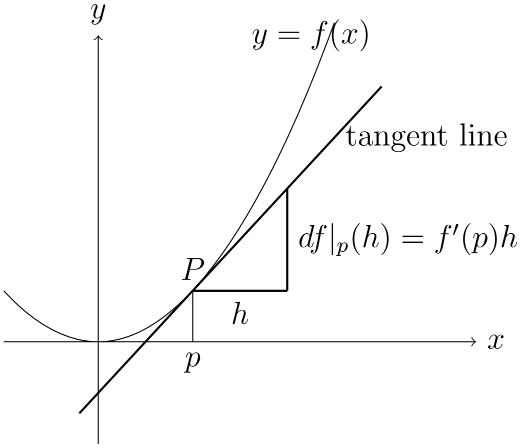


Figure 30.1: The differential df at p

The *differential* of f *at* the point p is the function which takes as input any horizontal step h and outputs the corresponding vertical rise

$$f'(p)h$$

if une follows the tangent line to $y = f(x)$ at $P(p, f(p))$. The differential of f at p is denoted $df|_p$ and its value *on* any $h \in \mathbb{R}$ is thus

$$df|_p(h) = f'(p)h. \tag{30.3}$$

For example, for the function sin we have

$$d\sin|_{\pi}(3) = \cos(\pi) * 3 = -3.$$

The function x, that is the function whose graph is $y = x$, has the value p at any point p:

$$x(p) = p \quad \text{for all } p \in \mathbb{R}.$$

The slope of $y = x$ is 1 everywhere and so

$$dx|_p(h) = 1 * h = h.$$

A *differential form* Φ is a function that associates to each point p in some domain set $S \subset \mathbb{R}$ a *linear* function $\Phi|_p$ on $\mathbb{R}$, that is

$$\Phi|_p(h) = K_p h \qquad \text{for all } h \in \mathbb{R},$$

for some $K_p \in \mathbb{R}$.

Thus, for a differentiable function f on some set $S \subset \mathbb{R}$, the differential

$$df$$

is a differential form.

The *product* of a differential form Φ with a function g is defined as the function whose value *at* p *on* h is

$$(g\Phi)|_p(h) = g(p)\Phi|_p(h). \tag{30.4}$$

Sometimes we write $g\Phi$ as Φg; thus

$$\Phi g = g\Phi \tag{30.5}$$

Working with a differentiable function f we have

$$(df(x))|_p(h) = f'(p)h$$

and this is exactly the same as

$$\big(f'(x)dx\big)|_p(h) = f'(p)h.$$

Hence we have the extremely useful notational identity:

$$df(x) = f'(x)dx. \tag{30.6}$$

All of this notation has been designed to produce this notational consistency:

$$\frac{df(x)}{dx} = f'(x), \tag{30.7}$$

where on the left we now have a genuine ratio (of functions), not just a formal one.

Using equation (30.6) we can easily verify the following convenient identities:

$$\begin{aligned} d(f+g) &= df + dg \\ dC &= 0 \qquad \text{if } C \text{ is constant} \\ d(fg) &= (df)g + f\,dg \\ d\left(\frac{f}{g}\right) &= \frac{g\,df - f\,dg}{g^2} \\ df(g(x)) &= f'(g(x))dg(x) \qquad \text{(this is from the chain rule),} \end{aligned} \tag{30.8}$$

where f and g are differentiable functions on some common domain except that in the last identity we assume the composite $f(g(x))$ is defined on some open interval. (Note $\frac{\Phi}{f}$ means $\frac{1}{f}\Phi$, for any differential form Φ and function f.)

As an example, we have

$$d\log(\sin x^2) = \frac{1}{\sin x^2}\cos(x^2) * 2x\,dx.$$

If f is a differentiable function on an interval containing points a and b we define the integral of the differential df to be

$$\int_a^b df = f(b) - f(a). \tag{30.9}$$

For example,

$$\int_{\pi/2}^{\pi} d(\sin x) = \sin\pi - \sin(\pi/2) = 0 - 1 = -1,$$

and

$$\int_1^0 de^x = e^0 - e^1 = 1 - e,$$

where note that the upper endpoint is actually $<$ than the lower endpoint.

All of what we have done in this section is essentially just notation. Now there is a nice convergence of notation with the Riemann integral: for any differentiable function g on any interval $[a, b]$, with $a, b \in \mathbb{R}$ and $a \leq b$, we have

$$\int_a^b g' = g(b) - g(a) = \int_a^b dg(x) = \int_a^b g'(x)\,dx \tag{30.10}$$

Thus,

$$\int_a^b g'(x)\,dx = \int_a^b g'. \tag{30.11}$$

If f is any continuous function on $[a, b]$ then by the fundamental theorem of calculus there is a function F on $[a, b]$ for which

$$F' = f.$$

Hence

$$\int_a^b f = \int_a^b F' = \int_a^b F'(x)\,dx = \int_a^b f(x)\,dx.$$

Thus we have, finally, a complete identification of the Riemann integral as the integral of a differential form:

$$\int_a^b f = \int_a^b f(x)\,dx. \tag{30.12}$$

The Riemann integral is rooted in ideas going back to Archimedes' computation of areas. The notation of differentials was invented in the 1600s by Leibniz, but a precise development of differential forms (in higher dimensions) was done only in the early 20th century by Élie Cartan. The equality (30.12) is, for us, a theorem (an easy consequence of the fundamental theorem of calculus, as we have seen) but traditionally (30.12) is viewed simply as different notation for the same integral.

30.3 Using the fundamental theorem

Let us start with a simple example. We will work out the integral

$$\int_0^1 x\,dx.$$

By Theorem 30.1.2, if we can find a function g for which $g'(x) = x$ on $[0, 1]$ then we can easily determine the value of the integral $\int_0^1 x\,dx$ as $g(1) - g(1)$. We can easily find a function g for which

$$g'(x) = x$$

for all $x \in \mathbb{R}$. Just recall that

$$(x^2)' = 2x,$$

and so

$$\frac{d(x^2/2)}{dx} = \frac{1}{2}2x = x.$$

Then by Theorem 30.1.2 we have

$$\int_0^1 x\,dx = \int_0^1 (x^2/2)'\,dx = \frac{x^2}{2}\Big|_0^1 = \frac{1^2}{2} - \frac{0^2}{2} = \frac{1}{2}.$$

Geometrically, this is the area under the graph of

$$y = x$$

over $x \in [0, 1]$. This is just a rght angled triangle with base 1 and height 1, and so its area is indeed $(1/2)1 \cdot 1 = 1/2$.

Next consider the area under the parabola

$$y = x^2$$

over $x \in [a, b]$, where $a, b \in \mathbb{R}$ with $a \leq b$. The area is

$$\int_a^b x^2\,dx.$$

We need to find a function whose derivative is x^2. Recall that

$$(x^3)' = 3x^2$$

and so

$$(x^3/3) = x^2.$$

Therefore

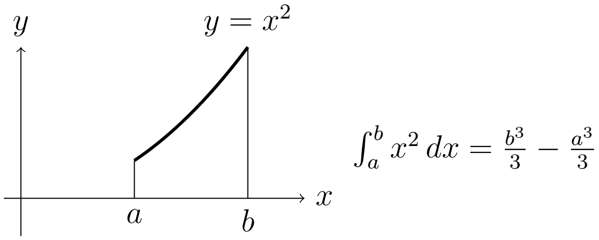


Figure 30.2: Area below $y = x^2$ for $x \in [a, b]$

$$\int_a^b x^2\,dx = \int_a^b (x^3/3)'\,dx \overset{\text{Thm.30.1.2}}{=} \frac{x^3}{3}\Big|_a^b = \frac{b^3}{3} - \frac{a^3}{3}.$$

Archimedes amazing determination of areas assocated with parabolas has thus been reduced to a simple routine calculation.

Next consider the area under

$$y = \sin x \qquad x \in [0, \pi].$$

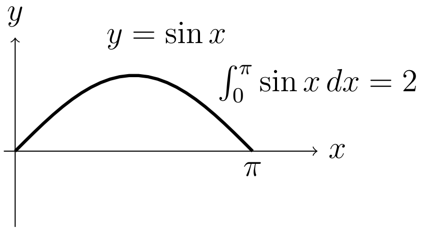


Figure 30.3: Area below $y = \sin x$ for $x \in [0, \pi]$

The area is

$$\begin{aligned} \int_0^\pi \sin x\,dx &= \int_0^\pi (-\cos x)'\,dx \\ &= -\cos x\Big|_0^\pi \\ &= (-\cos \pi) - (-\cos 0) = (-(-1)) - (-1) \\ &= 2. \end{aligned} \tag{30.13}$$

30.4 Indefinite integrals

With the examples of the previous section, it is clear that the task of finding a a function g whose derivative g' is a given function f is crucial to computing integrals exactly. The *indefinite integral* of a function f is the general function g for which

$$g' = f.$$

The indefinite integral is denoted

$$\int f(x)\,dx,$$

or with t or some other 'dummy variable' in place of x.

For example, we know that

$$(x^3/3)' = x^2.$$

If $g(x)$ is any function whose derivative is also x^2 then

$$\frac{d\left(g(x) - \frac{x^3}{3}\right)}{dx} = g'(x) - x^2 = 0,$$

and, assuming we are working over $\mathbb{R}$ (or any interval), it follows that

$$g(x) - \frac{x^3}{3} = \text{constant}.$$

Denoting this 'arbitrary constant' by C we have

$$g(x) = \frac{x^3}{3} + C.$$

Thus we have the indefinite integral of x^2:

$$\int x^2\,dx = \frac{x^3}{3} + C,$$

where C is an arbitrary constant. The presence of this arbitrary constant ensures that we have the 'general' solution to the problem of finding a function whose derivative is x^2, not just one special choice. Of course, since the difference between two such solutions is just a constant, it is not a matter of great importance.

In general, in mathematics it is not good practice to use ambiguous notation, but writing

$$\int f(x)\,dx$$

for not one function but for the class of all functions with derivative $f(x)$ is worth the occasional discomfort caused by the ambiguity.

If a differential form Phi is expressed as

$$\Phi = f(x)\,dx$$

then we have the integral of Φ

$$\int \Phi \overset{\text{def}}{=} \int f(x)\,dx.$$

Thus an integral of a differential form Φ is a function g for which

$$dg = \Phi.$$

Such integrals are of great interest and use in higher dimensions.

For any integer n, *except for* $n = -1$,

$$\int x^n\,dx = \frac{x^{n+1}}{n+1} + C, \tag{30.14}$$

with C being an arbitrary constant, as you can readily check by differentiating $x^{n+1}/(n+1)$:

$$\left(x^{n+1}/(n+1)\ \ + C\right)' = \frac{(n+1)x^n}{n+1} = x^n.$$

What if $n = -1$? Then we are seeking a function $g(x)$ whose derivative is $1/x$:

$$g'(x) = \frac{1}{x}.$$

Recall that

$$\log' t = \frac{1}{t} \qquad \text{for } t \in (0, \infty).$$

Thus we can write

$$\int \frac{1}{x}\,dx = \log x + C, \tag{30.15}$$

for any arbitary constant C, provided we restrict the functions all to the positive ray $(0,\infty)$.

Is there a function defined for negative values of x with derivative $1/x$? It is easily checked that

$$\frac{d\log(-x)}{dx} = [\log'(-x)]\cdot(-1) = \frac{1}{-x}(-1) = \frac{1}{x} \quad \text{for } x \in (-\infty, 0).$$

Thus sometimes it is convenient to use the combined formula

$$\log|x| + C$$

for the indefinite integral of $1/x$. However, $\log x$ and $\log(-x)$ do not really splice together naturally, and not simply because of the necessary gap at $x = 0$. A full exploration of this would require going into the complex plane, a subject well beyond our objectives.

30.5 Revisiting the exponential function

Historically the logarithm log was invented before the exponential function. However, mathematically, it seems more natural to develop the exponential function e^x first and then define log as the inverse function. This is essentially what we have done, except that we never really defined e^x in a logically connected way. In section 26.5 we devloped the properties of e^x by *assuming* that there is a function exp on $\mathbb{R}$ with the following two properties:

$$\begin{aligned} \exp' &= \exp \\ \exp(0) &= 1. \end{aligned} \tag{30.16}$$

Using this assumption we proved that

$$\exp(x) = e^x \qquad \text{for all } x \in \mathbb{R},$$

where

$$e = \exp(1).$$

The function exp has a strictly positive derivative, which implies that it has an inverse function. We defined log to be the inverse function of exp, defining $\log a$ to be the unique real number for which

$$e^{\log a} = a,$$

with $\log a$ being defined for all $a \in (0, \infty)$. Then we showed that

$$\log' x = \frac{1}{x} \qquad \text{for all } x \in (0, \infty).$$

Having developed integration theory we are not at a point where we can turn this logical development on its head (thereby regaining the correct, if strange, historical development) by *defining* the function log directly as

$$\log a = \int_1^a \frac{1}{x}\, dx \qquad \text{for all } a \in (0, \infty). \tag{30.17}$$

Since $1/x$ is continuous on $(0, \infty)$, the fundamental theorem of calculus guarantees that the integral defining $\log a$ exists and is finite; moreover it also assures us that

$$\log' a = \frac{1}{a} \qquad \text{for all } a \in (0, \infty). \tag{30.18}$$

This derivative being strictly positive, log has an inverse function, call it exp. Its derivative is (by Proposition 24.1.1):

$$\exp'(y) = \frac{1}{\log' x},$$

where $y = \log x$, and so

$$\exp'(y) = \frac{1}{1/x} = x = \exp(y).$$

Moreover, the definition of exp as inverse of log along with the simple fact that

$$\log(1) = 0$$

shows that

$$\exp(0) = 1.$$

Thus exp is indeed a function satisfying both conditions in (30.16), making the development of both exp and log logically complete.

Chapter 31

Riemann Sum Examples

In this chapter we work out some Riemann sums and compare them with the corresponding integrals.

31.1 Riemann sums for $\int_1^N \frac{dx}{x^2}$

Let N be a an integer > 1. Consider the partition of $[1, N]$ given by

$$P = \{1, 2, \ldots, N\}.$$

This breaks up $[1, N]$ into $N - 1$ intervals, each of width 1:

$$[1, 2],\ [2, 3]\ \ldots, [N - 1, N].$$

The k-th interval has the form

$$[k, k + 1].$$

We will work out the upper and lower Riemann sums for $1/x^2$ relative to P. On the interval

$$[k, k + 1]$$

the highest value of $1/x^2$ is $1/k^2$ and the lowest value is $1/(k+1)^2$. The area of the k-th lower rectangle is therefore

$$\text{area of } k\text{-th lower rectangle} = \frac{1}{(k + 1)^2} \cdot 1,$$

because its width is 1. Similarly

$$\text{area of } k\text{-th upper rectangle} = \frac{1}{k^2} \cdot 1.$$

Hence the lower sum is

$$L(1/x^2, P) = \sum_{k=1}^{N-1} \frac{1}{(k+1)^2} \cdot 1 = \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{N^2}.$$

Similarly, the upper sum is

$$U(1/x^2, P) = \sum_{k=1}^{N-1} \frac{1}{k^2} \cdot 1 = \frac{1}{1^2} + \frac{1}{2^2} + \cdots + \frac{1}{(N-1)^2}.$$

The actual integral

$$\int_1^N \frac{1}{x^2}\, dx$$

lies between these value:

$$L(1/x^2, P) \leq \int_1^N \frac{1}{x^2}\, dx \leq U(1/x^2, P). \tag{31.1}$$

The upper and lower sums are not really very good approximations to the value of the intergal because they are separated by quite a bit:

$$U(1/x^2, P) - L(1/x^2, P) = \frac{1}{1^2} - \frac{1}{N^2} = 1 - \frac{1}{N^2},$$

which tends to 1 as $N \to \infty$.

We can work out the integral of $1/x^2$ using the fundamental theorem of calculus:

$$\int_1^N \frac{dx}{x^2} = \int_1^N d\left(-\frac{1}{x}\right) = -\frac{1}{N} - \left(-\frac{1}{1}\right) = 1 - \frac{1}{N}.$$

Using this in (31.1) produces:

$$\frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{N^2} \leq 1 - \frac{1}{N} \leq \frac{1}{1^2} + \frac{1}{2^2} + \cdots + \frac{1}{(N-1)^2}. \tag{31.2}$$

We can extract some information from this by focusing on the first inequality:

$$s_N \overset{\text{def}}{=} \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{N^2} \leq 1 - \frac{1}{N}. \tag{31.3}$$

This is true for all integers $N \in \{2, 3, \ldots\}$. Observe that the sequence of sums s_1, s_2, ... increases in value as additional terms are added on:

$$s_1 < s_2 < s_3 < \cdots$$

Therefore, there is a limit

$$\lim_{N\to\infty} s_N = \sup_{N\in\{2,3,\ldots\}} s_N.$$

From (31.3) we see that

$$\lim_{N\to\infty} \leq 1.$$

Thus,

$$\lim_{N\to\infty} \left[\frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{N^2}\right] \leq 1.$$

This limit is displayed as an 'infinite series sum':

$$\frac{1}{2^2} + \frac{1}{3^2} + \cdots$$

Since the value of this is finite (being ≤ 1) the value of

$$\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots$$

is also finite, having value ≤ 2. One says that the series

$$\sum_n \frac{1}{n^2} = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots \tag{31.4}$$

converges.

The convergence of the series above can be seen in other ways, but the method using the integral $\int 1/x^2\,dx$ is useful for other similar sums as well.

The actual value of the sum (31.4) can be computed by more advanced methods; the amazing result is

$$\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots = \frac{\pi^2}{6}. \tag{31.5}$$

This identity was established by Euler in 1735.

Observe that

$$\lim_{t\to\infty}\int_1^t \frac{dx}{x^2} = \lim_{t\to\infty}\left[1-\frac{1}{t}\right] = 1.$$

We can interpret this to mean that the area under the graph

$$y = \frac{1}{x^2}$$

all the way over $[1,\infty)$ is 1:

$$\int_1^\infty \frac{dx}{x^2} = 1, \tag{31.6}$$

where we are taking the integral $\int_1^\infty$ on the left side to mean the limit of $\int_1^t$ as $t \to \infty$.

31.2 Riemann sums for $1/x$

For the partition

$$P = \{1, 2, \ldots, N\}$$

of $[1, N]$, where N is any integer > 1, and the function $1/x$ we have the lower sum

$$L(1/x, P) = \frac{1}{2}\cdot 1 + \cdots + \frac{1}{N}\cdot 1$$

and the upper sum

$$U(1/x, P) = \frac{1}{1}\cdot 1 + \frac{1}{2}\cdot 1 + \cdots + \frac{1}{N-1}\cdot 1.$$

The exact integral, which lies between these sums, is

$$\int_1^N \frac{dx}{x} = \log(x)\Big|_1^N = \log(N) - \log(1) = \log(N).$$

Hence

$$\frac{1}{2}\cdot 1 + \cdots + \frac{1}{N}\cdot 1 \leq \log N \leq \frac{1}{1}\cdot 1 + \frac{1}{2}\cdot 1 + \cdots + \frac{1}{N-1}\cdot 1. \tag{31.7}$$

Unlike what happened with $1/x^2$, we have an infinite area under the graph of $1/x$ over $[1,\infty)$:

$$\int_1^\infty \frac{dx}{x} \overset{\text{def}}{=} \lim_{t\to\infty} \int_1^t \frac{dx}{x} = \lim_{t\to\infty} \log(t) = \infty. \tag{31.8}$$

Looking back to the second inequality in (31.7) we have

$$\lim_{N\to\infty} \left[\frac{1}{1}+\frac{1}{2}+\cdots\right] \geq \lim_{N\to\infty} \log(N) = \infty,$$

and so

$$\sum_{n=1}^{\infty} \frac{1}{n} = \frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\cdots = \infty. \tag{31.9}$$

The series $\sum_{n=1}^{\infty}$ is called the *harmonic series* and the fact that the sum is ∞ is expressed by saying that series is *divergent*.

The difference between the upper sum and the integral over $[1,N]$ is

$$\sum_{k=1}^{N-1} \frac{1}{k} - \log(N).$$

It turns out that this has a finite limit as $N\to\infty$:

$$\gamma = \lim_{N\to\infty} \left[\frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{N} - \log(N)\right], \tag{31.10}$$

called *Euler's constant.*

31.3 Riemann sums for x

We focus on the function given by $f(x)=x$ on $[0,1]$.

Let N be a an integer >1. Consider the partition of $[0,1]$ given by

$$P = \left\{0, \frac{1}{N}, \ldots, \frac{N}{N}\right\}.$$

This breaks up $[0,1]$ into N intervals, each of width 1:

$$\left[0,\frac{1}{N}\right], \left[\frac{1}{N},\frac{2}{N}\right] \ldots, \left[\frac{N-1}{N},\frac{N}{N}\right].$$

The k-th interval has the form

$$\left[\frac{k-1}{N}, \frac{k}{N}\right].$$

We will work out the upper and lower Riemann sums for the function x relative to P. On the interval

$$\left[\frac{k-1}{N}, \frac{k}{N}\right]$$

the sup of x is just $\frac{k}{N}$ and the inf is $\frac{k-1}{N}$.

The area of the k-th lower rectangle is therefore

$$\text{area of } k\text{-th lower rectangle} = \frac{k-1}{N} \cdot \frac{1}{N}$$

because its width is 1. Similarly

$$\text{area of } k\text{-th upper rectangle} = \frac{k}{N} \cdot \frac{1}{N}.$$

Hence the lower sum is

$$\begin{aligned} L(x,P) &= \sum_{k=1}^{N} \frac{k-1}{N} \cdot \frac{1}{N} \\ &= \frac{1}{N}\left[0 + \frac{1}{N} + \frac{2}{N} + \cdots + \frac{N-1}{N}\right] \qquad (31.11) \\ &= \frac{1}{N^2}\left[1 + 2 + \cdots + (N-1)\right]. \end{aligned}$$

Similarly, the upper sum is

$$\begin{aligned} U(x,P) &= \sum_{k=1}^{N} \frac{k}{N} \cdot \frac{1}{N} \\ &= \frac{1}{N}\left[\frac{1}{N} + \frac{2}{N} + \cdots + \frac{N}{N}\right] \qquad (31.12) \\ &= \frac{1}{N^2}\left[1 + 2 + \cdots + N\right]. \end{aligned}$$

The integral

$$\int_0^1 x\,dx$$

lies between these value:

$$L(x,P) \le \int_0^1 x\,dx \le U(x,P). \tag{31.13}$$

Observe that $U(x,P)$ differs from $L(x,P)$ by the term $\frac{1}{N^2}N$:

$$U(x,P) - L(x,P) = \frac{1}{N} \to 0 \qquad \text{as } N \to \infty.$$

This implies that there can be only one number lying between the upper sums and the lower sums and hence that the integral

$$\int_0^1 x\,dx$$

exists and

$$\int_0^1 x\,dx = \lim_{N\to\infty} \frac{1}{N^2}\left[1+2+\cdots+N\right].$$

Now we use the sum formula (see (31.24) below):

$$1+2+\cdots+N = \frac{N(N+1)}{2}. \tag{31.14}$$

Hence

$$\int_0^1 x\,dx = \lim_{N\to\infty} \frac{1}{N^2}\frac{N(N+1)}{2} = \lim_{N\to\infty} \frac{1}{N^2}\frac{N\cdot N(1+1/N)}{2} = \lim_{N\to\infty} \frac{(1+1/N)}{2}.$$

Hence

$$\int_0^1 x\,dx = \frac{1}{2}. \tag{31.15}$$

This result is, of course, far easier to see using the fundamental theorem of calculus:

$$\int_0^1 x\,dx = \int_0^1 \left(\frac{1}{2}x^2\right)' dx = \left.\frac{x^2}{2}\right|_0^1 = \frac{1}{2}. \tag{31.16}$$

31.4 Riemann sums for x^2

We focus on the function given by $f(x) = x^2$ on $[0, 1]$.

Let N be a an integer > 1. We work with the partition of $[0, 1]$ given by

$$P = \left\{0, \frac{1}{N}, \ldots, \frac{N}{N}\right\}.$$

The k-th interval marked off by this partition is

$$\left[\frac{k-1}{N}, \frac{k}{N}\right].$$

On the interval

$$\left[\frac{k-1}{N}, \frac{k}{N}\right]$$

the sup of x^2 is just $\frac{k^2}{N^2}$ and the inf is $\frac{(k-1)^2}{N^2}$.

The area of the k-th lower rectangle is therefore

$$\text{area of } k\text{-th lower rectangle} = \frac{(k-1)^2}{N^2} \cdot \frac{1}{N}$$

because its width is 1. Similarly

$$\text{area of } k\text{-th upper rectangle} = \frac{k^2}{N^2} \cdot \frac{1}{N}.$$

Hence the lower sum is

$$\begin{aligned} L(x^2, P) &= \sum_{k=1}^{N} \frac{(k-1)^2}{N^2} \cdot \frac{1}{N} \\ &= \frac{1}{N}\left[\frac{0^2}{N^2} + \frac{1^2}{N^2} + \frac{2^2}{N^2} + \cdots + \frac{(N-1)^2}{N^2}\right] \qquad (31.17) \\ &= \frac{1}{N^3}\left[1^2 + 2^2 + \cdots + (N-1)^2\right]. \end{aligned}$$

Similarly, the upper sum is

$$\begin{aligned} U(x^2, P) &= \sum_{k=1}^{N} \frac{k^2}{N^2} \cdot \frac{1}{N} \\ &= \frac{1}{N}\left[\frac{1^2}{N^2} + \frac{2^2}{N^2} + \cdots + \frac{N^2}{N^2}\right] \qquad (31.18) \\ &= \frac{1}{N^3}\left[1^2 + 2^2 + \cdots + N^2\right]. \end{aligned}$$

The integral

$$\int_0^1 x^2\,dx$$

lies between these value:

$$L(x^2,P) \le \int_0^1 x^2\,dx \le U(x^2,P). \tag{31.19}$$

The upper sum $U(x^2,P)$ differs from the lower sum $L(x^2,P)$ by $\frac{1}{N^3}N^2$:

$$U(x^2,P) - L(x^2,P) = \frac{1}{N} \to 0 \qquad \text{as } N \to \infty.$$

Hence there can be only one number lying between the upper sums and the lower sums and so the integral

$$\int_0^1 x^2\,dx$$

exists and

$$\int_0^1 x^2\,dx = \lim_{N\to\infty} \frac{1}{N^3}\left[1^2 + 2^2 + \cdots + N^2\right].$$

We now need the sum formula (see (31.31) below):

$$1^2 + 2^2 + \cdots + N^2 = \frac{N(N+1)(2N+1)}{6}. \tag{31.20}$$

Hence

$$\begin{aligned}\int_0^1 x^2\,dx = \lim_{N\to\infty} \frac{1}{N^3}\frac{N(N+1)(2N+1)}{6} &= \lim_{N\to\infty} \frac{1}{N^3}\frac{N\cdot N(1+1/N)N(2+1/N)}{6}\\ &= \lim_{N\to\infty} \frac{(1+1/N)(2+1/N)}{6}\\ &= \frac{1}{3}.\end{aligned} \tag{31.21}$$

Hence

$$\int_0^1 x^2\,dx = \frac{1}{3}. \tag{31.22}$$

This, of course, agrees with the result using the fundamental theorem of calculus:

$$\int_0^1 x^2\,dx = \int_0^1 \left(\frac{1}{3}x^3\right)' dx = \frac{x^3}{3}\Big|_0^1 = \frac{1}{3}. \tag{31.23}$$

31.5 Power sums

We will work out a few power sum formulas, starting with

$$1+2+\cdots+N=\frac{N(N+1)}{2}. \tag{31.24}$$

There are several ways of proving this formula. Let

$$S_1=1+2+\cdots+N.$$

Writing the same sum backwards we have:

$$\begin{aligned} S_1 &= 1\ +\ 2\ +\cdots+(N-1)\ +\ N \\ S_1 &= N\ +\ (N-1)\ +\cdots+2\ +\ 11 \end{aligned} \tag{31.25}$$

In the first equation above the terms on the right increase by 1 at each step whereas in the second equation the terms decrease by 1 at each stage. Adding them we have

$$2S_1=(N+1)\ +(N+1)\ +\cdots+\ (N+1)=N(N+1).$$

Hence

$$S_1=\frac{N(N+1)}{2}.$$

Here is a second, longer and far less appealing method. We start with

$$(a+1)^2=a^2+2a+1$$

which implies

$$(a+1)^2-a^2=2a+1.$$

Now use $a=1$, through $a+N$ in this to obtain:

$$\begin{aligned} 2^2-1^2 &= 2*1\ +\ 1 \\ 3^2-2^2 &= 2*2\ +\ 1 \\ \vdots\vdots\vdots & \ \vdots \\ (N+1)^2-N^2 &= 2*N+1 \end{aligned} \tag{31.26}$$

Adding all these up we see that on the left all terms cancel except for the very first one -1^2 and the very last one $(N+1)^2$:

$$(N+1)^2-1^1=2*\underbrace{[1+2+\cdots+N]}_{S_1}+N*1.$$

Hence

$$2S_1 = (N+1)^2 - 1 - N = (N+1)^2 - (N+1) = (N+1-1)(N+1) = N(N+1),$$

which implies

$$S_1 = \frac{N(N+1)}{2}. \tag{31.27}$$

The advantage of this method is that it works for sums of higher powers. We use this method to find the sum

$$S_2 = 1^2 + 2^2 + \cdots + N^2. \tag{31.28}$$

We start now with

$$(a+1)^3 = a^3 + 3a^2 + 3a + 1$$

from which we have

$$(a+1)^3 - a^3 = 3a^2 + 3a + 1.$$

Using $a = 1$ through $a = N$ in this produces

$$\begin{aligned} 2^3 - 1^3 &= 3 * 1^2 \ + 3 * 1 \ + \ 1 \\ 3^3 - 2^3 &= 3 * 2^2 \ + 3 * 2 \ + \ 1 \\ \vdots\vdots\vdots\vdots & \ \vdots \\ (N+1)^3 - N^3 &= 3 * N^2 \ + 3 * N \ + \ 1. \end{aligned} \tag{31.29}$$

When we add these up on the left everything cancels except the very first term -1^3 and the very last term $(N+1)^3$, and we have:

$$(N+1)^3 - 1^3 = 3 * \underbrace{\left[1^2 + 2^2 + \cdots + N^2\right]}_{S_2} + 3 \underbrace{\left[1 + 2 + \cdots + N\right]}_{S_1} + N * 1.$$

Hence

$$3S_2 = (N+1)^3 - 1 - 3S_1 - N$$

Using the formula (31.27) we have

$$\begin{aligned}
3S_2 &= (N+1)^3 - 1 - N - 3S_1\\
&= (N+1)^3 - (N+1) - 3\frac{N(N+1)}{2}\\
&= (N+1)\left[(N+1)^2 - 1 - \frac{3N}{2}\right] = (N+1)\left[N^2+2N+1-1-\frac{3N}{2}\right]\\
&= (N+1)\left[N^2+\frac{N}{2}\right] = (N+1)N[N+\frac{1}{2}]\\
&= (N+1)N\frac{(2N+1)}{2}.
\end{aligned} \tag{31.30}$$

Hence

$$S_2 = \frac{N(N+1)(2N+1)}{6}. \tag{31.31}$$

Applied to the sum of the cubes this strategy produces

$$1^3 + 2^3 + \cdots + N^3 = \left[\frac{N(N+1)}{2}\right]^2. \tag{31.32}$$

Amazingly, this is just the square of the sum S_1:

$$1^3 + 2^3 + \cdots + N^3 = [1 + 2 + \cdots + N]^2. \tag{31.33}$$

Chapter 32

Integration Techniques

In this chapter we study some techniques for working out indefinite integrals. What this means is that we are given a function $f(x)$ and we have to find a function $g(x)$ for which

$$dg(x) = f(x)dx.$$

Since $dg(x)$ is $g'(x)dx$ this means finding a function g whose derivative if f:

$$g'(x) = f(x).$$

We write the general such function as the *indefinite integral*

$$g(x) = \int f(x)\, dx.$$

Here $f(x)$ is called the *integrand*.

32.1 Substitutions

For the integral

$$\int (x+1)^5\, dx$$

it would be a long and slow method to work out the fifth power and then integrate. Instead we substitute y for $x+1$ and rewrite the integral in terms of y:

$$y = x + 1$$

Then

$$dy = 1\,dx$$

and so

$$\int (x+1)^5\,dx = \int y^5\,dy = \frac{1}{6}y^6 + C = \frac{1}{6}(x+1)^6 + C,$$

where C is an arbitrary constant.

The essence of the idea behind the substitution method is simple. We inspect the integral

$$\int f(x)\,dx$$

and write it in the form

$$\int F\bigl(p(x)\bigr)\,p'(x)dx,$$

for some functions F and p, and then substitute

$$y = p(x)$$

to transform the given integral as

$$\int f(x)\,dx = \int F\bigl(p(x)\bigr)\,p'(x)dx = \int F\bigl(p(x)\bigr)\,dp(x) = \int F(y)\,dy,$$

and, if al goes well, the integral $\int F(y)\,dy$ is 'simpler' than what we started with, thereby reducing $\int f(x)\,dx$ to a simpler integral. The main challenge is in identifying the functions F and p which express $f(x)$ as $F(p(x))p'(x)$.

As we will see below there are also some simple variations on this strategy. For example, it may be easier to write $f(x)$ as a constant multiple of $F(p(x))p'(x)$ or, in some cases, we can break up $f(x)$ into a sum of pieces, each of which is easier to work out separately.

Consider

$$\int (2x-5)^{3/5}\,dx.$$

We substitute

$$z = 2x - 5$$

which gives

$$dz = 2dx \qquad \text{and so} \qquad dx = \frac{1}{2}dz$$

and so

$$\begin{aligned}\int (2x-5)^{3/5}\,dx = \int z^{3/5}\,\frac{1}{2}dz &= \frac{1}{2}\int z^{\frac{3}{5}}\,dz\\ &= \frac{1}{2} * \frac{1}{\frac{3}{5}+1} z^{\frac{3}{5}+1} + C\\ &= \frac{5}{16}(2x-5)^{8/5} + C,\end{aligned}$$

where C is an arbitrary constant.

For

$$\int (4-3x)^{2/7}\,dx$$

the substitution would be

$$y = 4-3x$$

which leads to

$$dy = -3dx \qquad \text{and so} \quad dx = -\frac{1}{3}dy$$

which then transforms the given integral as follows:

$$\begin{aligned}\int (4-3x)^{2/7}\,dx = \int y^{2/7}\left(-\frac{1}{3}dy\right) &= -\frac{1}{3}\int y^{2/7}\,dy\\ &= -\frac{1}{3} * \frac{1}{\frac{2}{7}+1} y^{\frac{2}{7}+1} + C\\ &= -\frac{7}{27}(4-3x)^{\frac{2}{7}+1} + C,\end{aligned}$$

where C is an arbitrary constant.

Sometimes the integrand should be reworked a bit before or after the substitution is made. For example, for

$$\int x(3-4x)^{2/5}\,dx$$

we can substitute

$$y = 3-4x \tag{32.1}$$

for which

$$dy = -4dx \qquad \text{so that} \qquad dx = -\frac{1}{4}dy,$$

and then the given integral has the form

$$\int xy^{2/5}\left(-\frac{1}{4}dy\right).$$

We need to replace the x in the integrand with its expression in terms of y by solving (32.1):

$$x = \frac{1}{4}(3-y).$$

Then our integral becomes

$$\int \frac{1}{4}(3-y)y^{2/5}\left(-\frac{1}{4}dy\right) = -\frac{1}{16}\int (3-y)y^{2/5}\,dy.$$

The right side looks complicated but can be worked out by breaking up into pieces:

$$\int (3-y)y^{2/5}\,dy = 3\int y^{2/5}\,dy - \int \underbrace{y\cdot y^{2/5}}_{y^{7/5}}\,dy = 3\frac{1}{\frac{2}{5}+1}y^{\frac{2}{5}+1} - \frac{1}{\frac{7}{5}+1}y^{7/5+1} + \text{constant}$$

Now putting everything together we have

$$\int x(3-4x)^{2/5}\,dx = \left(-\frac{1}{16}\right)\frac{15}{7}y^{7/5} - \left(-\frac{1}{16}\right)\frac{5}{12}y^{12/5} + C,$$

where C is an arbitrary constant and y is as in (32.1). Thus

$$\int x(3-4x)^{2/5}\,dx = -\frac{15}{112}(3-4x)^{2/5} + \frac{5}{192}(3-4x)^{12/5} + C,$$

for any arbitrary constant C.

Moving on to more complicated integrands, consider

$$\int (x^2+1)^{2/3}x\,dx$$

Observe that xdx is about the same as $d(x^2+1)$, aside from a constant multiple. The substitution is

$$y = x^2 + 1.$$

With this we have

$$dy = 2xdx \quad \text{and so} \quad xdx = \frac{1}{2}dy$$

which converts the given integral as follows:

$$\int (x^2+1)^{2/3} xdx = \int y^{2/3}\frac{1}{2}dy = \frac{1}{2}\int y^{2/3}\,dy = \frac{1}{2}\frac{1}{\frac{2}{3}+1}y^{\frac{2}{3}+1} + C,$$

where C is any constant. Rewriting in terms of x we have

$$\int (x^2+1)^{2/3} x\,dx = \frac{3}{10}(x^2+1)^{5/3} + C.$$

Here is a faster display of this method:

$$\begin{aligned}\int (x^2+1)^{2/3} x\,dx &= \int (x^2+1)^{\frac{2}{3}}\frac{1}{2}d(x^2+1)\\ &= \frac{1}{2}\frac{1}{\frac{2}{3}+1}(x^2+1)^{2/3+1} + C\\ &= \frac{3}{10}(x^2+1)^{5/3} + C.\end{aligned}$$

The same strategy works for more complicated integrands. For example, for

$$\int (3x^2 - 12x + 2)^{-7/3}(x-2)\,dx$$

observe again that $(x-2)dx$ is a constant multiple of $d(3x^2-5x+2)$; so we use

$$y = 3x^2 - 12x + 2$$

for which

$$dy = (6x-12)\,dx = 6(x-2)dx$$

and this transforms the given integral as follows:

$$\int (3x^2-12x+2)^{-7/3}(x-2)\,dx = \int y^{-7/3}\frac{1}{6}dy = \frac{1}{6}\int y^{-7/3}\,dy,$$

which integrates out to

$$\int (3x^2-12x+2)^{-7/3}(x-2)\,dx = \frac{1}{6}\frac{1}{-\frac{7}{3}+1}y^{-\frac{7}{3}+1}+C = -\frac{1}{8}(3x^2-12x+2)^{-\frac{4}{3}}+C,$$

for any arbitrary constant C.

Sometimes the substitution is not as obvious as in the preceding (manufactured) example. Consider

$$\int \frac{x}{\sqrt{x^2+1}}\,dx.$$

Observe that the numerator is x, which is constant times the dervative of x^2+1. So we use

$$u = x^2+1$$

for which

$$du = 2xdx$$

and so

$$\int \frac{x}{\sqrt{x^2+1}}\,dx = \int \frac{\frac{1}{2}du}{\sqrt{u}} = \int \frac{1}{2\sqrt{u}}\,du = \sqrt{u}+C,$$

where C is an arbitrary constant; thus,

$$\int \frac{x}{\sqrt{x^2+1}}\,dx = \sqrt{x^2+1}+C. \tag{32.2}$$

Similarly,

$$\begin{aligned} \int \frac{x}{\sqrt{3-2x^2}}\,dx &= \int \frac{-\frac{1}{4}d(3-2x^2)}{\sqrt{3-2x^2}}\,dx \\ &= -\frac{1}{2}\int \frac{du}{2\sqrt{u}} \qquad \text{with } u = 3-2x \\ &= -\frac{1}{2}\sqrt{u}+C \\ &= -\frac{1}{2}\sqrt{3-2x^2}+C, \end{aligned} \tag{32.3}$$

where C is an arbitrary constant.

Substitutions need not be limited to polynomials or algebraic functions. For

$$\int \frac{dx}{x\log x}$$

we observe the x in the denominator and recall that $d\log x = 1/x$; so we rewrite the integral as

$$\int \frac{dx}{x\log x} = \int \frac{\frac{1}{x}dx}{\log x} = \int \frac{d(\log x)}{\log x} = \log(\log x)+C,$$

for any arbitrary constant C.

Similarly,

$$\int \frac{dx}{x(\log x)^{2/3}} = \int \frac{d(\log x)}{(\log x)^{2/3}} = \int (\log x)^{-2/3}\, d(\log x) = \frac{1}{-\frac{2}{3}+1}(\log x)^{-\frac{2}{3}+1} + C,$$

where C is the arbitrary constant.

Here is an example with trigonometric functions:

$$\int \sin x \cos x \, dx = \int \sin x \, d(\sin x) = \frac{1}{2}\sin^2 x + C, \tag{32.4}$$

where C is an arbitrary constant. We can do the same integral using the substitution $\cos x$ instead:

$$\int \sin x \cos x \, dx = \int \cos x \sin x \, dx = -\int \cos x \, d(\cos x) = -\frac{1}{2}\cos^2 x + C_1, \tag{32.5}$$

where C_1 is an abitrary constant. For the first time we have now denoted the arbitrary constant by C_1 instead of C and this is the first time we face a possible pitfall of the ambiguity in the meaning of the indefinite integral. Observe that

$$\sin^2 x = 1 - \cos^2 x$$

transforms the first expression (32.4) for $\int \sin x \cos x \, dx$ into

$$\frac{1}{2}(1 - \cos^2 x) + C = -\frac{1}{2}\cos^2 x + \frac{1}{2} + C,$$

and this agrees with (32.5) on taking C_1 to be $C + 1/2$.

Here is a slightly more involved example of this type of substitution:

$$\int \sin^6 x \cos x \, dx = \int \sin^6 x \, d(\sin x) = \frac{1}{7}\sin^7 x + C, \tag{32.6}$$

where C is an arbitrary constant.

We can take this a step beyond, with

$$\int \sin^4 x \cos^7 x \, dx.$$

We write this as

$$\int \sin^4 x \cos^6 x \; \cos x \, dx = \int \sin^4 x \cos^6 x \, d(\sin x).$$

Now we make a key observation: the 6-th power $\cos^6 x$ can be written as

$$\cos^6 x = (\cos^2 x)^3 = (1 - \sin^2 x)^3.$$

This brings us to

$$\int \sin^4 x \cos^7 x \, dx = \int \sin^4 x (1 - \sin^2 x)^3 \, d(\sin x).$$

Now we can use the substitution

$$y = \sin x$$

to transform the given integral to

$$\int y^4 (1 - y^2)^3 \, dy.$$

Here we have the integral of a polynomial and though it is a lengthy write-up the integration is routine:

$$\begin{aligned}\int y^4(1-y^2)^3 \, dy &= \int y^4(1 - 3y^2 + 3y^4 - y^6) \, dy \\ &= \int (y^4 - 3y^6 + 3y^8 - y^{10}) \, dy \\ &= \frac{1}{5}y^5 - \frac{3}{7}y^7 + \frac{3}{9}y^9 - \frac{1}{11}y^{11} + C,\end{aligned}$$

where C is the arbitrary constant; substituting back $\sin x$ for y produced the complete integral:

$$\int \sin^4 x \cos^7 x \, dx = \frac{1}{5}\sin^5 x - \frac{3}{7}\sin^7 x + \frac{3}{9}\sin^9 x - \frac{1}{11}\sin^{11} x + C.$$

32.2 Some trigonometric integrals

As starting point we have the integrals

$$\begin{aligned}\int \sin x \, dx &= -\cos x + C_1 \\ \int \cos x \, dx &= \sin x + C_2\end{aligned} \tag{32.7}$$

where C_1 and C_2 are arbitrary constants.

Next up we have the simplest substitutions:

$$\int \sin(3x)\,dx = \int \sin(3x)\,\left(\frac{1}{3}d(3x)\right) = \frac{1}{3}\int \sin(3x)\,d(3x) = -\frac{1}{3}\cos(3x) + C$$

and, similarly,

$$\int \cos(5x)\,dx = \frac{1}{5}\sin(5x) + C',$$

with C and C' being arbitrary constants.

Our next objective is to work out integrals of the form

$$\int \sin(Ax)\cos(Bx), dx$$

and other such integrals of products of trigonometric functions.

The key strategy is the use of the trigonometric sum formulas such as

$$\begin{aligned} \sin(a+b) &= \sin a \cos b + \sin b \cos a \\ \sin(a-b) &= \sin a \cos b - \sin b \cos a \end{aligned} \tag{32.8}$$

Adding these we get

$$\sin(a+b) + \sin(a-b) = 2\sin a \sin b,$$

and so

$$\sin a \cos b = \frac{1}{2}\left[\sin(a+b) + \sin(a-b)\right]. \tag{32.9}$$

Thus

$$\sin(3x)\cos(7x) = \frac{1}{2}\left[\sin(10x) + \sin(-4x)\right] = \frac{1}{2}\left[\sin(10x) - \sin(4x)\right].$$

Integrating, we obtain

$$\int \sin(3x)\cos(7x)\,dx = \frac{1}{2}\left[-\frac{1}{10}\cos(10x) + \frac{1}{4}\cos(4x)\right] + C,$$

where C is any constant.

The sum formula method works for other trigonometric products. For these, recall

$$\begin{aligned} \cos(a+b) &= \cos a \cos b - \sin a \sin b \\ \cos(a-b) &= \cos a \cos b + \sin a \sin b. \end{aligned} \tag{32.10}$$

Adding we get

$$\cos(a+b) + \cos(a-b) = 2\cos(a)\cos(b),$$

and subtracting we get

$$\cos(a-b) - \cos(a+b) = 2\sin a \sin(b).$$

Hence

$$\begin{aligned}\cos a \cos b &= \frac{1}{2}\left[\cos(a+b) + \cos(a-b)\right] \\ \sin a \sin b &= \frac{1}{2}\left[\cos(a-b) - \cos(a+b)\right].\end{aligned} \tag{32.11}$$

Using these we have

$$\begin{aligned}\int \cos(3x)\cos(5x)\,dx &= \frac{1}{2}\int \left[\cos(8x) + \cos(2x)\right]\,dx \quad \text{(note that } \cos(-2x) = \cos(2x)\text{)} \\ &= \frac{1}{2}\left[\frac{1}{8}\sin(8x) + \frac{1}{2}\sin(2x)\right] + C\end{aligned} \tag{32.12}$$

for any constant C.

Similarly,

$$\begin{aligned}\int \sin(2x)\sin(6x)\,dx &= \frac{1}{2}\int \left[\cos(4x) - \cos(8x)\right]\,dx \quad \text{(note that } \cos(-4x) = \cos(4x)\text{)} \\ &= \frac{1}{2}\left[\frac{1}{4}\sin(4x) + \frac{1}{8}\sin(8x)\right] + C\end{aligned} \tag{32.13}$$

for any constant C.

We can apply this even to

$$\int \sin^2 x\,dx,$$

viewing the integrand as the product $\sin x \sin x$:

$$\sin^2 x = \sin x \sin x = \frac{1}{2}\left[\cos(0) - \cos(2x)\right] = \frac{1}{2}\left[1 - \cos(2x)\right],$$

from which we have

$$\begin{aligned}\int \sin^2 x\,dx &= \frac{1}{2}\int [1-\cos(2x)]\,dx \\ &= \frac{1}{2}\left[x-\frac{1}{2}\sin(2x)\right] + \text{constant.} \\ &= \frac{x}{2}-\frac{1}{4}\sin(2x)+\frac{1}{2}+C,\end{aligned} \tag{32.14}$$

where C is an arbitrary constant. Using the formula

$$\sin(2x) = \sin x \cos x$$

we an rewrite the above integral also as:

$$\int \sin^2 x\,dx = \frac{x}{2}-\frac{1}{2}\sin x\cos x + C.$$

Similarly for $\cos^2 x$ we have

$$\cos^2 x = \cos x \cos x = \frac{1}{2}[\cos(0)+\cos(2x)] = \frac{1}{2}[1+\cos(2x)],$$

which leads to

$$\int \cos^2 x\,dx = \frac{1}{2}\left[x+\frac{1}{2}\sin(2x)\right]. \tag{32.15}$$

We can use the sum-formula strategy multiple times. For example,

$$\begin{aligned}\sin(5x)\sin(3x)\cos(4x) &= \frac{1}{2}[\cos(2x)-\cos(8x)]\cos(4x) \\ &= \frac{1}{2}[\cos(2x)\cos(4x)-\cos(8x)\cos(4x)] \\ &= \frac{1}{2}\left[\frac{1}{2}[\cos(6x)+\cos(2x)]-\frac{1}{2}[\cos(12x)+\cos(4x)]\right] \\ &= \frac{1}{4}[\cos(6x)+\cos(2x)-\cos(12x)-\cos(4x)]\end{aligned} \tag{32.16}$$

from which we have

$$\int \sin(5x)\sin(3x)\cos(4x)\,dx = \frac{1}{4}\left[\frac{1}{6}\sin(6x)+\frac{1}{2}\sin(2x)-\frac{1}{12}\sin(12x)-\frac{1}{4}\sin(4x)\right]$$

32.3 Summary of basic trigonometric integrals

Let us recall the following derivatives

$$\begin{aligned}\sin' x &= \cos x\\ \cos' x &= -\sin x\\ \tan' x &= \sec^2 x\\ \csc' x &= -\csc x \cot x\\ \sec' x &= \sec x \tan x\\ \cot' x &= -\csc^2 x.\end{aligned} \tag{32.17}$$

These invert to give the following integrals:

$$\begin{aligned}\int \cos x\, dx &= \sin x + C\\ \int \sin x\, dx &= -\cos x + C\\ \int \sec^2 x\, dx &= \tan x\\ \int \csc x \cot x\, dx &= -\csc x\\ \int \sec x \tan x\, dx &= \sec x + C\\ \int \csc^2 x.\, dx &= -\cot x + C.\end{aligned} \tag{32.18}$$

Some natural entries are missing from this list. For instance,

$$\tan x\, dx.$$

This integral can be worked out by observing that in

$$\int \tan x\, dx = \int \frac{\sin x}{\cos x}\, dx,$$

the numerator $\sin x$ is minus the derivative of the denominator $\cos x$. So we use the substitution

$$y = \cos x$$

which gives

$$dy = -\sin x\,dx$$

and so

$$\tan x\,dx = \int \frac{-dy}{y} = -\int \frac{dy}{y} = -\log y + C,$$

and so

$$\int \tan x\,dx = -\log\cos x + C.$$

Since

$$-\log A = \log\frac{1}{A},$$

we can also write the integral as

$$\int \tan x\,dx = \log(\sec x) + C. \tag{32.19}$$

If we are working over an interval on which $\sec x$ is negative we should use the absolute value form:

$$\int \tan x\,dx = \log(\sec x) + C.$$

(In complex analysis the correct integral is (32.19).)

The integral

$$\int \sec x\,dx$$

is also of interest. There is no natural method for this integral that leads to the value in its simplest form. The following tricky method is based on already happening on the correct answer by accident:

$$\begin{aligned}\int \sec x\,dx &= \int \frac{\sec x(\sec x + \tan x)}{\sec x + \tan x}\,dx\\ &= \int \frac{\sec^2 x + \sec x\tan x}{\sec x + \tan x}\,dx\end{aligned} \tag{32.20}$$

and here we make the remarkable observation that the numerator is the derivative of the denominator:

$$(\sec x + \tan x)' = \sec x\tan x + \sec^2 x = \sec x(\sec x + \tan x),$$

and so we use the subtitution

$$y = \sec x + \tan x.$$

Then, as we just observed,

$$dy = (\sec x) y \, dx,$$

and so

$$\sec x \, dx = \frac{dy}{y},$$

from which we have

$$\int \sec x \, dx = \int \frac{dy}{y} = \log y + C.$$

Thus

$$\int \sec x \, dx = \log(\sec x + \tan x) + C. \tag{32.21}$$

Again, we can use $\log |\cdots|$ if $\sec x + \tan x$ is negative.

32.4 Using trigonometric substitutions

For the integral

$$\int \frac{dx}{\sqrt{1 - x^2}}$$

the best substitution is

$$x = \sin \theta.$$

This means we are setting θ to be an inverse sin of x; for definiteness we can set

$$\theta = \sin^{-1}(x),$$

which restricts θ to $[-\pi/2, \pi/2]$. Then

$$\begin{aligned}\int \frac{dx}{\sqrt{1-x^2}} &= \int \frac{\cos(\theta)\,d\theta}{\sqrt{1-\sin^2\theta}} \\ &= \int \frac{\cos(\theta)\,d\theta}{\sqrt{\cos^2\theta}} \qquad \text{(which shows why we use } x = \sin\theta\text{)} \\ &= \int \frac{\cos\theta\,d\theta}{\cos\theta} \qquad \text{(here we use } \cos\theta \geq 0 \text{ for } \theta \in [-\pi/2, \pi/2].\text{)} \\ &= \int d\theta \\ &= \theta + C \\ &= \sin^{-1} x + C,\end{aligned} \tag{32.22}$$

where C is an arbitrary constant. We have seen before that

$$\frac{d\sin^{-1}(x)}{dx} = \frac{1}{\sqrt{1-x^2}},$$

which confirms the integration result.

A similar integral is

$$\int \frac{dx}{1+x^2}.$$

Here the best substitution is

$$x = \tan\theta$$

for which both

$$dx = \sec^2\theta\,d\theta$$

and

$$1 + x^2 = 1 + \tan^2\theta = \sec^2\theta,$$

which simplied the integral:

$$\int \frac{dx}{1+x^2} = \int \frac{\sec^2\theta\,d\theta}{\sec^2\theta} = \int \theta = \theta + C = \tan^{-1} x + C,$$

for any arbitrary constant C.

More involved is

$$\int \sqrt{1-x^2}\,dx.$$

When we substitute

$$x = \sin\theta$$

and

$$dx = \cos\theta\, d\theta$$

we obtain

$$\int \sqrt{1-x^2}\, dx = \int \sqrt{1-\sin^2\theta}\, \cos\theta\, d\theta = \int \sqrt{\cos^2\theta}\, \cos\theta\, d\theta \tag{32.23}$$

and here again we have the annoying fact that (at least working with real numbers)

$$\sqrt{\cos^2\theta} = |\cos\theta|.$$

We get past this bit of unpleasantness by requiring that

$$\theta = \sin^{-1} x \in [-\pi/2, \pi/2],$$

which ensures that

$$\cos\theta \geq 0$$

and so

$$\sqrt{\cos^2\theta} = \cos\theta.$$

Returning to the integration (32.23) we have

$$\int \sqrt{1-x^2}\, dx = \int \cos\theta\, \cos\theta\, d\theta = \int \cos^2\theta\, d\theta.$$

Looking back at (32.15) we have then

$$\int \sqrt{1-x^2}\, dx = \frac{1}{2}\left[\theta + \frac{1}{2}\sin(2\theta)\right] + C,$$

where C is an arbitrary constant. We have to substitute back in $\theta = \sin^{-1} x$. Before doing this observe that

$$\sin(2\theta) = 2\sin\theta\cos\theta$$

and so, since

$$\sin\theta = x \quad \text{and} \quad \cos\theta = \sqrt{1-x^2},$$

we conclude that

$$\int \sqrt{1-x^2}\,dx = \frac{1}{2}\left[\sin^{-1} x + x\sqrt{1-x^2}\right] + C. \tag{32.24}$$

Sometimes we need to do some algebra before using a trigonometric substitution. Consider for example

$$\int \sqrt{-4x^2+12x-6}\,dx.$$

It is best to work the term inside the $\sqrt{\cdots}$ into a 'completed squares' form; for convenience we work with the negative so that the coefficient of x^2 is positive:

$$\begin{aligned} 4x^2-12x+6 &= (2x)^2 - 2*(2x)*3+3^2 \quad -3^2+6 \\ &= (2x-3)^2+[6-3^2] \quad (\text{using } (A-B)^2 = A^2-2AB+B^2) \\ &= (2x-3)^2-3. \end{aligned} \tag{32.25}$$

Thus

$$-4x^2+12x-6 = 3-(2x-3)^2,$$

and so

$$\int \sqrt{-4x^2+12x-6}\,dx = \int \sqrt{3-(2x-3)^2}\,dx.$$

Once we have this form we need to observe that this looks roughly like (32.23). We could now use a sin substitution. Alternatively, we can make the similarity with (32.23) greater by substituting

$$2x-3 = \sqrt{3}y, \tag{32.26}$$

which ensures that

$$-4x^2+12x-6 = 3-(2x-3)^2 = 3-3y^2 = 3(1-y^2), \tag{32.27}$$

and

$$2\,dx = \sqrt{3}\,dy.$$

So

$$\int \sqrt{-4x^2+12x-6}\,dx = \int \sqrt{3(1-y^2)}\,\frac{\sqrt{3}}{2}\,dy = \frac{\sqrt{3}}{2}\sqrt{3}\int \sqrt{1-y^2}\,dy.$$

Using (32.24) we have then

$$\int \sqrt{-4x^2+12x-6}\,dx = \frac{3}{2}\frac{1}{2}\left[\sin^{-1} y + y\sqrt{1-y^2}\right] + C.$$

Substituting in the value of y from (32.26) yields the complete answer. The algebra can be made nicer by observing that

$$\begin{aligned}\frac{3}{2}\frac{1}{2}\left[\sin^{-1} y + y\sqrt{1-y^2}\right] &= \frac{1}{4}\left[\sin^{-1} y + 3y\sqrt{1-y^2}\right]\\ &= \frac{1}{4}\left[\sin^{-1} y + \sqrt{3}y\sqrt{3(1-y^2)}\right]\\ &= \frac{1}{4}\left[\sin^{-1}\left(\frac{2x-3}{\sqrt{3}}\right) + (2x-3)\sqrt{-4x^2+12x-6}\right]\end{aligned} \tag{32.28}$$

where in the final step we used (32.27) and the value of y from (32.26).

32.5 Integration by parts

The product rule for derivatives is

$$d(UV) = UdV + VdU,$$

and so

$$\int d(UV) = \int UdV + \int VdU,$$

provided we line up the 'arbitrary' constants appropriately. Then

$$\int U\,dV = UV - \int V\,dU.$$

This often helps in simplifying integrals, and is called the *integration by parts* formula or method.

For example,

$$\begin{aligned}\int \log x\,dx &= (\log x)x - \int x\,d(\log x)\\ &= x\log x - \int x\frac{dx}{x}\\ &= x\log x - \int dx\\ &= x\log x - x + C.\end{aligned} \tag{32.29}$$

Thus,

$$\int \log x\, dx = x \log x - x + C, \tag{32.30}$$

where C is an arbitrary constant.

Sometimes the choice of U and V requires some planning ahead:

$$\begin{aligned}
\int x \log x\, dx &= \frac{1}{2}\int \log x\, d(x^2) \\
&= \frac{1}{2}(\log x)x^2 - \frac{1}{2}\int x^2\, d(\log x) \\
&= \frac{1}{2}(\log x)x^2 - \frac{1}{2}\int x^2 \frac{dx}{x} \\
&= \frac{1}{2}(\log x)x^2 - \frac{1}{2}\int x\, dx \\
&= \frac{1}{2}(\log x)x^2 - \frac{1}{6}x^3 + C,
\end{aligned} \tag{32.31}$$

where C is an arbitrary constant.

We can apply this method to

$$\int x \sin x\, dx$$

as follows:

$$\begin{aligned}
\int x \sin x\, dx &= -\int x\, d(\cos x) \\
&= -\left[x(\cos x) - \int \cos x\, dx\right] \\
&= -x \cos x + \int \cos x\, dx \\
&= -x \cos x + \sin x + C
\end{aligned} \tag{32.32}$$

where C is an arbitrary constant.

Here is a considerable more involved use of integration by parts:

$$\begin{aligned}\int \sec^3 x\,dx &= \int \sec x \sec^2 x\,dx\\ &= \int \sec x\,d(\tan x)\\ &= \sec x\tan x - \int \tan x\,d(\sec x)\\ &= \sec x\tan x - \int \tan x\ \sec x\tan x\,dx\\ &= \sec x\tan x - \int \sec x\tan^2 x\,dx\\ &= \sec x\tan x - \int \sec x(\sec^2 x - 1)\,dx \quad (\text{using } \sec^2 x = \tan^2 x + 1)\\ &= \sec x\tan x - \int \sec^3\ dx + \int \sec x\,dx\end{aligned} \tag{32.33}$$

and at this stage it looks bad at first because we have ended up with $\int \sec^3 x\,dx$ again, on the right. But we are saved because there is a minus sign in front of the integral on the right; we can move it to the left to get

$$2\int \sec^3 x\,dx = \sec x\tan x + \int \sec x\,dx = \sec x\tan x + \log(\sec x + \tan x) + \text{constant},$$

on using formula (32.21) for $\int \sec x\,dx$. Hence,

$$\int \sec^3 x\,dx = \frac{1}{2}\left[\sec x\tan x + \log(\sec x + \tan x)\right] + C, \tag{32.34}$$

where C is an arbitrary constant. As usual, a real solution is also obtained if $\sec x + \tan x$ on the right is negative by replacing this with its absolute value.

One other example of this type is

$$\int e^{Ax}\sin(Bx)\,dx$$

where A and B are non-zero constants. Then we have

$$\begin{aligned}\int e^{Ax}\sin(Bx)\,dx &= -\frac{1}{B}\int e^{Ax}\,d\left(\cos(Bx)\right)\\ &= -\frac{1}{B}\left[e^{Ax}\cos(Bx)-\int\cos(Bx)\,d(e^{Ax})\right]\\ &= -\frac{1}{B}\left[e^{Ax}\cos(Bx)-\int\cos(Bx)\,Ae^{Ax}\,dx\right] \qquad (32.35)\\ &= -\frac{1}{B}\left[e^{Ax}\cos(Bx)-A\int\cos(Bx)\,e^{Ax}\,dx\right]\\ &= -\frac{1}{B}e^{Ax}\cos(Bx)+\frac{A}{B}\int\cos(Bx)\,e^{Ax}\,dx.\end{aligned}$$

Now we run the same method on the integral $\int\cos(Bx)\,e^{Ax}\,dx$:

$$\begin{aligned}\int e^{Ax}\sin(Bx)\,dx &= -\frac{1}{B}e^{Ax}\cos(Bx)+\frac{A}{B}\int e^{Ax}\left(\frac{1}{B}d\sin(Bx)\right)\\ &= -\frac{1}{B}e^{Ax}\cos(Bx)+\frac{A}{B^2}\int e^{Ax}\,d\sin(Bx)\\ &= -\frac{1}{B}e^{Ax}\cos(Bx)+\frac{A}{B^2}\left[e^{Ax}\sin(Bx)-\int\sin(Bx)de^{Ax}\right]\\ &= -\frac{1}{B}e^{Ax}\cos(Bx)+\frac{A}{B^2}\left[e^{Ax}\sin(Bx)-\int\sin(Bx)Ae^{Ax}\,dx\right]\\ &= -\frac{1}{B}e^{Ax}\cos(Bx)+\frac{A}{B^2}e^{Ax}\sin(Bx)-\frac{A^2}{B^2}\int e^{Ax}\sin(Bx)\,dx,\end{aligned}$$
$$(32.36)$$

with our original integral reappearing on the right side, with a negative sign. Keeping in mind that the integrals on the two sides might differ by a constant we have:

$$\left(1+\frac{A^2}{B^2}\right)\int e^{Ax}\sin(Bx)\,dx = -\frac{1}{B}e^{Ax}\cos(Bx)+\frac{A}{B^2}e^{Ax}\sin(Bx)+\text{constant}.$$

Multiplying both sides by B^2 gives

$$\begin{aligned}(A^2+B^2)\int e^{Ax}\sin(Bx)\,dx &= -Be^{Ax}\cos(Bx)+Ae^{Ax}\sin(Bx)+\text{constant}\\ &= e^{Ax}\left[\cos(Bx)+A\sin(Bx)-B\cos(Bx)\right]+\text{constant}\end{aligned}$$
$$(32.37)$$

Dividing by $A^2 + B^2$ produces, at last, the formula

$$\int e^{Ax} \sin(Bx)\, dx = e^{Ax} \left[\frac{A \sin(Bx) - B \cos(Bx)}{A^2 + B^2} \right] + C, \tag{32.38}$$

where C is an arbitrary constant. We assumed A and B are both nonzero. We can check easily that the formula works even if one of these two values is 0.

Exercises on Integration by Substitution

1. Work out the following integrals using substitutions:
    (a) $\int (4 - 3x)^{2/3}\, dx$
    (b) $\int \sqrt{2 + 5x}\, dx$
    (c) $\int \frac{1}{\sqrt{2-3x}}\, dx$
    (d) $\int x(3 - 2x)^{4/5}\, dx$
    (e) $\int \frac{x}{(2+5x)^{3/5}}\, dx$
    (f) $\int \frac{2x+1}{\sqrt{x^2+x+5}}\, dx$
    (g) $\int e^{-x^2/2} x\, dx$
    (h) $\int \frac{\sqrt{\log(2x+5)}}{2x+5}\, dx$
    (i) $\int \frac{1}{x \log(x) \log(\log x)}\, dx$
    (j) $\int \sin(5x) \cos(2x)\, dx$
    (k) $\int \sin(5x) \sin(2x)\, dx$
    (l) $\int \cos(5x) \cos(2x)\, dx$
    (m) $\int \sin^3 x\, dx$
    (n) $\int \cos^3 x\, dx$
    (o) $\int \sin^2(5x)\, dx$
    (p) $\int \sqrt{3 - 6x - x^2}\, dx$
    (q) $\int \frac{1}{\sqrt{3-6x-x^2}}\, dx$

Chapter 33

Paths and Length

33.1 Paths

A *path* c in the plane $\mathbb{R}^2$ is a mapping

$$c : I \to \mathbb{R}^2 : t \mapsto c(t) = \bigl(xc(t), yc(t)\bigr), \tag{33.1}$$

where I is some interval in $\mathbb{R}$. We can think of $c(t)$ as being the *position* of a point at *time* t.

In (33.1) we are denoting the x-coordinate of a point p by $x(p)$:

$$x(p) = x\text{-coordinate of a point } p,$$

so that the x-coordinate of $c(t)$ is $x\bigl(c(t)\bigr)$, which we write briefly as

$$xc(t).$$

Similarly, the y-coordinate of a point p is

$$y(p) = y\text{-coordinate of a point } p,$$

and the y-coordinate of $c(t)$ is $y\bigl(c(t)\bigr)$, which we write briefly as $yc(t)$.

As our first example, consider

$$c(t) = (t, 2t + 1) \qquad \text{for } t \in \mathbb{R}.$$

Think of this as a moving point, whose position at clock time t is $(t, 2t + 1)$; see Figure 33.1.

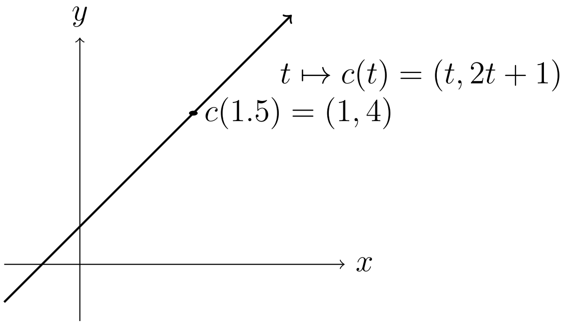


Figure 33.1: The path $c : \mathbb{R} \to \mathbb{R}^2 : t \mapsto (t, 2t+1)$

This is a point traveling at a uniform speed along a straight line. How fast is it traveling? We can check how fast the x and y coordinates are changing:

$$c'(t) = \big((xc)'(t), (yc)'(t)\big) = \left(\frac{dt}{dt}, \frac{d(2t+1)}{dt}\right) = (1, 2).$$

This is called the *velocity* of the path c at time t. Note that for this path the velocity is the same, being $(1, 2)$, at all times t.

Here is a *different* path that also travels along the same line, but with increasing speed:

$$\mathbb{R} \to \mathbb{R}^2 : t \mapsto c(t) = (t^2, 2t^2 + 1).$$

This is displayed in Figure 33.2.

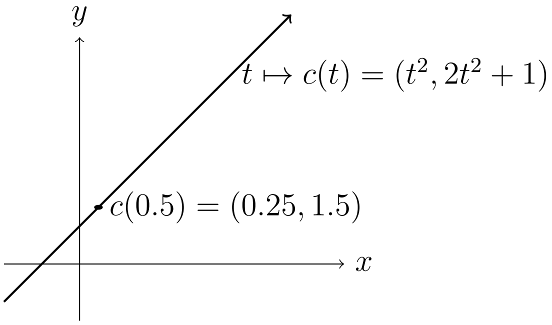


Figure 33.2: The path $c : \mathbb{R} \to \mathbb{R}^2 : t^2 \mapsto (t, 2t^2+1)$

The velocity of this path at time t is

$$c'(t) = (2t, 4t),$$

which is clearly changing as the clock time t changes.

The path

$$[0, 2\pi] \to \mathbb{R}^2 : t \mapsto \bigl(\cos t, \sin t\bigr)$$

traces out the unit circle counterclockwise exactly once, starting out at

$$(\cos 0, \sin 0) = (1, 0)$$

and ending also at

$$(\cos 2\pi, \sin 2\pi) = (1, 0).$$

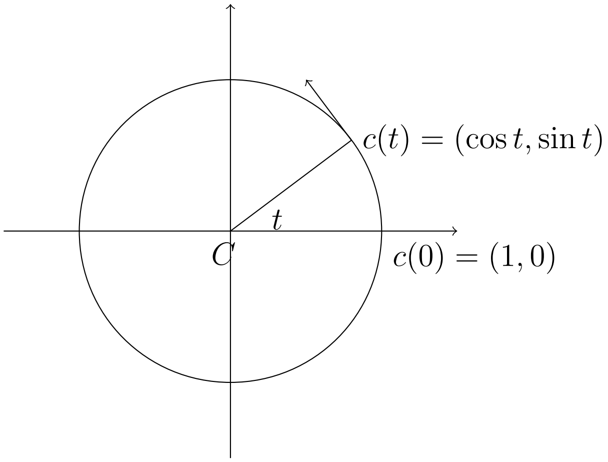


Figure 33.3: The path $[0, 2\pi] \to \mathbb{R}^2 : t \mapsto (\cos t, \sin t)$.

The velocity at time t is:

$$c'(t) = (-\sin t, \cos t).$$

In this example we have a different interpretation of t also possible: t is simply the measure of the angle between the x-axis and the line from the origin to the location of the point.

A path c is said to be *continuous* if its x and y components xc and yc are continuous. It is said to be differentiable at t if its x and y components are differentiable at t. The *velocity* of c at time t is

$$c'(t) = \bigl((xc)'(t), (yc)'(t)\bigr). \tag{33.2}$$

The *speed* of the path c at time t is defined to be

$$|c'(t)| = \sqrt{(xc)'(t)^2 + (yc)'(t)^2} \tag{33.3}$$

33.2 Lengths of paths

Consider a path

$$c : [a, b] \to \mathbb{R}^2,$$

where $a, b \in \mathbb{R}$ and $a < b$. Let

$$P = \{t_0, t_1, \ldots, t_N\}$$

be a partition of $[a, b]$, with

$$a = t_0 < t_1 < \ldots < t_N = b.$$

We think of these as time instants when we 'observe' the path and note where it is. Suppose we approximate the path c by a path that travels in a straight line from the initial point $c(t_0)$ to the next point $c(t_1)$, and then straight to $c(t_2)$, and in this way till $c(t_N) = c(b)$. The *length* of this *polygonal* approximation is

$$l(c; P) = d\left(c(t_0), c(t_1)\right) + d\left(c(t_1), c(t_2)\right) + \ldots + d\left(c(t_{N-1}), c(t_N)\right), \tag{33.4}$$

where $d(P, Q)$ means the usual distance between points P and Q:

$$d(P, Q) = \sqrt{(x_Q - x_P)^2 + (y_Q - y_P)^2} \tag{33.5}$$

where

$$P = (x_P, y_P) \quad \text{and} \quad Q = (x_Q, y_Q).$$

If we refine the partition P by adding one more point, say s, which lies between t_{j-1} and t_j, then for the new partition P_1 we have a corresponding length

$$l(c; P_1).$$

The difference between this and $l(c; P)$ is clearly

$$l(c; P_1) - l(c; P) = d\left(c(t_{j-1}), c(s)\right) + d\left(c(s), c(t_j)\right) \quad - \quad d\left(c(t_{j-1}), c(t_j)\right),$$

and this is ≥ 0 because of the triangle inequality for distances:

$$d(P, S) + d(S, Q) \geq d(P, Q),$$

for any points P, S and Q.

Thus adding points to a partition increases the length of the corresponding polygonal approximation. So if P and P' are partitions of $[a, b]$ with

$$P \subset P',$$

we have

$$l(c; P) \leq l(c; P'). \tag{33.6}$$

If we keep on adding points to a partition, making it finer and finer, intuitively it is clear that the lengths of the polygonal approximations should approach the 'length' of the path c itself. Hence we define the *length* of c to be

$$l(c) = \sup_{\text{partitions } P \text{ of } [a,b]} l(c; P). \tag{33.7}$$

If c has a continuous derivative then we have a clean and simple formula for the length of c;

$$l(c) = \int_a^b \sqrt{(xc)'(t)^2 + (yc)'(t)^2}\, dt. \tag{33.8}$$

Let us apply this to the circle

$$c(t) = (\cos t, \sin 2) \quad t \in [0, 2\pi].$$

We have

$$l(c) = \int_0^{2\pi} \sqrt{(-\sin t)^2 + (\cos t)^2}\, dt = \int_0^{2\pi} 1\, dt = 2\pi,$$

confirming that the circumference of the unit circle is 2π.

33.3 Paths and Curves

The notion of a *curve* is natural but there are some choices available in defining what it is precisely. Briefly, a curve is a path but without worrying about the specific speed at which the path moves. Thus,

$$\mathbb{R} \to \mathbb{R}^2 : t \mapsto (t + 1, \sin(t + 1))$$

and

$$\mathbb{R} \to \mathbb{R}^2 : t \mapsto (t, \sin t)$$

are the same curve, the only difference between them being that the second path is always a bit 'ahead' of the first one.

On the other hand, the paths

$$\mathbb{R} \to \mathbb{R}^2 : t \mapsto (t, 2t)$$

and

$$\mathbb{R} \to \mathbb{R}^2 : (t^2, 2t^2)$$

don't correspond to the same curve because the second path always has x-coordinate nonnegative whereas the first one has negative values of x.

We say that a path

$$c_1 : [a, b] \to \mathbb{R}^2$$

is a *reparametrization* of a path

$$c_2 : [a', b'] \to \mathbb{R}^2$$

if

$$c_2(t) = c_1\bigl(\phi(t)\bigr) \qquad \text{for all } t \in [a', b'],$$

for some clock-changing mapping

$$\phi : [a', b'] \to [a, b] : t \mapsto \phi(t)$$

that is continuous and satisfies

$$\phi(a') = a \quad \text{and} \quad \phi(b') = b,$$

and

$$\phi(p) < \phi(q) \quad \text{if} \qquad p < q$$

for all $p, q \in [a, b]$. Thus the path c_2 is at time t where c_1 is at time $\phi(t)$.

We can say that two paths which are reparametrizations of each other correspond to the same *curve*. Alternatively, a *curve* is a path along with all of its reparametrizations.

Here is an important observation:

Proposition 33.3.1 *If a path c_1 is a reparametrization of a path c_2 then they have the same lengths:*

$$l(c_1) = l(c_2).$$

Thus we can speak of the 'length of a curve'. This is also called *arc length.*

There is some flexibility in defining curves. One could insist, for example, that only differentiable reparametrizations be allowed.

33.4 Lengths for graphs

Consider now a function

$$f : [a, b] \to \mathbb{R}$$

where $a, b \in \mathbb{R}$ with $a < b$. There is a corresponding natural path

$$[a, b] \to \mathbb{R}^2 : x \mapsto \big(x, f(x)\big),$$

which traces out the graph of f 'from left to right'. The length of this path is

$$\int_a^b \sqrt{1 + f'(x)^2}\, dx. \tag{33.9}$$

Here is an example. The graph of

$$x^{2/3} + y^{2/3} = 1 \tag{33.10}$$

had four parts, one in the positive quadrant and the others obtained by reflections across the two axes. We will work out the length of the full curve (a path that goes around the graph). This length is

$$4\int_0^1 \sqrt{1 + (y')^2}\, dx.$$

Now from the equation (33.10) we have on taking d/dx of both sides:

$$\frac{2}{3}x^{-1/3} + \frac{2}{3}y^{-1/3}y' = 0,$$

and so, on doing the algebra, we have

$$y' = -\frac{y^{1/3}}{x^{1/3}}.$$

Then

$$1 + (y')^2 = 1 + \frac{y^{2/3}}{x^{2/3}} = \frac{x^{2/3} + y^{2/3}}{x^{2/3}} = \frac{1}{x^{2/3}}$$

and so

$$\int_0^1 \sqrt{1 + (y')^2}\, dx = \int_0^1 \sqrt{1x^{2/3}}\, dx = \int_0^1 x^{-1/3}\, dx = \left.\frac{x^{-1/3+1}}{-\frac{1}{3}+1}\right|_0^1 = \frac{3}{2}.$$

So the length of the full curve is

$$4 \cdot \frac{3}{2} = 6.$$

Next consider the curve

$$y = x^2 \quad x \in [0, a],$$

for any $a > 0$. See Figure 33.4.

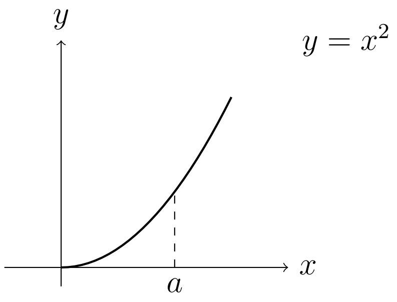


Figure 33.4: Arc length for $y = x^2$

The length is

$$\int_0^1 \sqrt{1 + (y')^2}\, dx = \int_0^a \sqrt{1 + 4x^2}\, dx$$

Recall that

$$\int \sqrt{1 + w^2}\, dw = \frac{1}{2}\left[w\sqrt{1 + w^2} + \log(w + \sqrt{1 + w^2}\right] + C,$$

where C is an arbitrary constant. Using

$$w = 2x$$

then gives

$$\int \sqrt{1 + 4x^2}\, 2dx = \frac{1}{2}\left[2x\sqrt{1 + 4x^2} + \log(2x + \sqrt{1 + 4x^2}\right] + C,$$

and so, dividing by 2, we have

$$\int \sqrt{1 + 4x^2}\, dx = \frac{1}{4}\left[2x\sqrt{1 + 4x^2} + \log(2x + \sqrt{1 + 4x^2}\right] + C,$$

for an arbitrary constant C. Hence

$$\int_0^a \sqrt{1+4x^2}\,dx = \frac{1}{4}\left[2a\sqrt{1+4a^2} + \log(2a + \sqrt{1+4a^2}\right].$$

Chapter 34

Selected Solutions

Solutions for Exercise Set 8.9.

1. $\lim_{x\to 1} 5 = 5$ because the sup's and inf's of the constant function 5 are both 5 on any neighborhood of 1.

2. $\lim_{x\to 1}(x^2 + 4x - \frac{5}{x}) = 1^2 + 4 \cdot 1 - \frac{5}{1} = 0$, by the rules for limits.

3. $\lim_{x\to 1} \frac{x^2-9}{x-3} = \frac{1^2-9}{1-3} = -8/(-2) = 4$.

4. $\lim_{x\to 1} \frac{x^3-1}{x^2-1}$

$$\begin{aligned}\lim_{x\to 1}\frac{x^3-1}{x^2-1} &= \lim_{x\to 1}\frac{(x-1)(x^2+x+1)}{(x-1)(x+1)}\\ &= \lim_{x\to 1}\frac{x^2+x+1}{x+1} = 3/2.\end{aligned}$$

5.

$$\begin{aligned}\lim_{x\to 1}\frac{x^4-1}{x^2-1} &= \lim_{x\to 1}\frac{(x-1)(x^3+x^2+x+1)}{(x-1)(x+1)}\\ &= \lim_{x\to 1}\frac{x^3+x^2+x+1}{x+1} = 4/2 = 2.\end{aligned}$$

6. $\lim_{x\to\infty} \frac{1}{x^2} = 0$.

7. $\lim_{x\to\infty} \frac{4x^3-3x+2}{x^2-x+1}$

$$\begin{aligned}\lim_{x\to\infty} \frac{4x^3-3x+2}{x^2-x+1} &= \lim_{x\to\infty} \frac{x^3\left(4-\frac{3}{x^2}+\frac{2}{x^3}\right)}{x^2\left(1-\frac{1}{x}+\frac{1}{x^2}\right)}\\ &= \lim_{x\to\infty} x\frac{\left(4-\frac{3}{x^2}+\frac{2}{x^3}\right)}{\left(1-\frac{1}{x}+\frac{1}{x^2}\right)}\\ &= \infty\cdot\frac{4-0+0}{1-0+0} = \infty.\end{aligned}$$

8.

$$\begin{aligned}\lim_{x\to\infty} \frac{5x^6-7x+2}{3x^6+x+2} &= \lim_{x\to\infty} \frac{x^6\left(5-\frac{7}{x^5}+\frac{2}{x^6}\right)}{x^6\left(3+\frac{1}{x^5}+\frac{2}{x^6}\right)}\\ &= \lim_{x\to\infty} \frac{5-\frac{7}{x^5}+\frac{2}{x^6}}{3+\frac{1}{x^5}+\frac{2}{x^6}}\\ &= \frac{5-0+0}{3+0+0} = 5/3.\end{aligned}$$

9.

$$\begin{aligned}\lim_{x\to\infty} \frac{4x^3+\sin x}{2x^3+\sqrt{x}} &= \lim_{x\to\infty} \frac{x^3\left(4+\frac{\sin x}{x^3}\right)}{x^3\left(2+\frac{\sqrt{x}}{x^3}\right)}\\ &= \lim_{x\to\infty} \frac{4+\frac{\sin x}{x^3}}{2+\frac{\sqrt{x}}{x^3}}\\ &= \frac{4+0}{2+0} = 2,\end{aligned}$$

where we used $(\sin x)/x^3 \to 0$ as $x \to \infty$ by using, for example, the ‘squeeze theorem’:

$$\left|\frac{\sin x}{x^3}\right| \le \left|\frac{1}{x^3}\right| \to 0 \quad \text{as } x \to \infty,$$

and also $\sqrt{x}/x^3 = \frac{1}{x^{3-1/2}} \to 0$ as $x \to \infty$.

10.

$$\begin{aligned}\lim_{x\to\infty}\frac{7x^5+x+\cos(x^3)}{2x^5-5x^2+1} &= \lim_{x\to\infty}\frac{x^5\left(7+\frac{1}{x^4}+\frac{\cos x^3}{x^5}\right)}{x^5\left(2-\frac{5}{x^3}+\frac{1}{x^5}\right)}\\ &= \lim_{x\to\infty}\frac{7+\frac{1}{x^4}+\frac{\cos x^3}{x^5}}{2-\frac{5}{x^3}+\frac{1}{x^5}}\\ &= \frac{7+0+0}{2-0+0}=\frac{7}{2},\end{aligned}$$

11. $\lim_{x\to\infty}\left[\sqrt{x+1}-\sqrt{x}\right]$

Sol: At first this looks like $\infty-\infty$ and that's bad. In order to deterime whether one of the two terms wins out over the other we need some trick. Here is a method that works often when square-roots are involved:

$$\begin{aligned}\lim_{x\to\infty}\left[\sqrt{x+1}-\sqrt{x}\right] &= \lim_{x\to\infty}\frac{\left[\sqrt{x+1}-\sqrt{x}\right]\left[\sqrt{x+1}+\sqrt{x}\right]}{\sqrt{x+1}+\sqrt{x}}\\ &= \lim_{x\to\infty}\frac{x+1\ -\ x}{\sqrt{x+1}+\sqrt{x}}\\ &\qquad\qquad\text{(using } (A-B)(A+B)=A^2-B^2\text{)}\\ &= \lim_{x\to\infty}\frac{1}{\sqrt{x+1}+\sqrt{x}}=0.\end{aligned}$$

12.

$$\begin{aligned}
\lim_{x\to\infty}\left[\sqrt{3x^2+1}-\sqrt{x^2+1}\right] &= \lim_{x\to\infty}\frac{\left[\sqrt{3x^2+1}-\sqrt{x^2+1}\right]\left[\sqrt{3x^2+1}+\sqrt{x^2+1}\right]}{\sqrt{3x^2+1}+\sqrt{x^2+1}}\\
&= \lim_{x\to\infty}\frac{3x^2+1\,-\,(x^2+1)}{\sqrt{3x^2+1}+\sqrt{x^2+1}}\\
&\qquad\qquad \text{(using } (A-B)(A+B)=A^2-B^2\text{)}\\
&= \lim_{x\to\infty}\frac{2x^2}{\sqrt{x^2\big(3+1/x^2\big)}+\sqrt{x^2(1+1/x^2)}}\\
&= \lim_{x\to\infty}\frac{2x^2}{x\sqrt{3+1/x^2}+x\sqrt{1+1/x^2}}\\
&= \lim_{x\to\infty}\frac{2x^2}{x\left[\sqrt{3+1/x^2}+\sqrt{1+1/x^2}\right]}\\
&= \lim_{x\to\infty}\frac{2x}{\sqrt{3+1/x^2}+\sqrt{1+1/x^2}}\\
&= \frac{\infty}{\sqrt{3}+1}\\
&= \infty.
\end{aligned}$$

13. $\lim_{x\to\infty}\left[\sqrt{4x^4+2}-\sqrt{x^4+1}\right]$

Use the same method as for the previous problem and reach

$$\begin{aligned}\lim_{x\to\infty}\left[\sqrt{4x^4+2}-\sqrt{x^4+1}\right] &= \text{stuff}\\ &= \lim_{x\to\infty}\frac{3x^4+1}{x^2\left[\sqrt{4+2/x^4}+\sqrt{1+1/x^4}\right]}\\ &= \lim_{x\to\infty}\frac{x^4(3+1/x^4)}{x^2\left[\sqrt{4+2/x^4}+\sqrt{1+1/x^4}\right]}\\ &= \lim_{x\to\infty}\frac{x^2(3+1/x^4)}{\sqrt{4+2/x^4}+\sqrt{1+1/x^4}}\\ &= \frac{\infty(3+0)}{2+1}=\infty.\end{aligned}$$

14. $\lim_{x\to\infty}\frac{\sqrt{x+1}}{\sqrt{x}} = \lim_{x\to\infty}\sqrt{\frac{x+1}{x}} = \lim_{x\to\infty}\sqrt{1+\frac{1}{x}} = \sqrt{1+0} = 1.$

15. $\lim_{x\to\infty} x\left[\sqrt{x^2+2}-\sqrt{x^2+1}\right]$

$$
\begin{aligned}
\lim_{x\to\infty} x\frac{\left[\sqrt{x^2+2}-\sqrt{x^2+1}\right]\left[\sqrt{x^2+2}+\sqrt{x^2+1}\right]}{\sqrt{x^2+2}+\sqrt{x^2+1}} &= \lim_{x\to\infty} x\frac{(x^2+2)-(x^2+1)}{\sqrt{x^2+2}+\sqrt{x^2+1}}\\
&= \lim_{x\to\infty} x\frac{1}{\sqrt{x^2(1+2/x^2)}+\sqrt{x^2(1+1/}}\\
&= \lim_{x\to\infty} x\frac{1}{x\sqrt{1+2/x^2}+x\sqrt{1+1/x^2)}}\\
&= \lim_{x\to\infty} x\frac{1}{x\left(\sqrt{1+2/x^2}+\sqrt{1+1/x^2)}\right)}\\
&= \lim_{x\to\infty} \frac{1}{\sqrt{1+2/x^2}+\sqrt{1+1/x^2}}\\
&= \frac{1}{1+1} = \frac{1}{2}.
\end{aligned}
$$

16.

$$
\begin{aligned}
\lim_{x\to\infty} \sqrt{x+2}\left[\sqrt{x+1}-\sqrt{x}\right] &= \lim_{x\to\infty} \sqrt{x+2}\frac{\left[\sqrt{x+1}-\sqrt{x}\right]\left[\sqrt{x+1}+\sqrt{x}\right]}{\left[\sqrt{x+1}+\sqrt{x}\right]} \\
&= \lim_{x\to\infty} \sqrt{x+2}\frac{x+1\ -\ x}{\sqrt{x+1}+\sqrt{x}} \\
&= \lim_{x\to\infty} \sqrt{x+2}\frac{1}{\sqrt{x+1}+\sqrt{x}} \\
&= \lim_{x\to\infty} \sqrt{x(1+2/x)}\frac{1}{\sqrt{x(1+1/x)}+\sqrt{x}} \\
\\
&= \lim_{x\to\infty} \sqrt{x}\sqrt{1+2/x}\frac{1}{\sqrt{x}\sqrt{1+1/x}+\sqrt{x}} \\
&= \lim_{x\to\infty} \sqrt{x}\sqrt{1+2/x}\frac{1}{\sqrt{x}\left(\sqrt{1+1/x}+1\right)} \\
&= \lim_{x\to\infty} \sqrt{1+2/x}\frac{1}{\sqrt{1+1/x}+1} \\
&= 1\cdot\frac{1}{1+1} = \frac{1}{2}.
\end{aligned}
$$

17. $\lim_{\theta\to 0} \frac{\sin(\theta^2)}{\theta^2} = \lim_{y\to 0} \frac{\sin y}{y} = 1$, on setting

$$y = \theta^2$$

and noting that $y \to 0$ as $\theta \to 0$.

18. $\lim_{\theta\to 0} \frac{\sin^2(\theta)}{\theta^2} = \lim_{\theta\to 0} \left(\frac{\sin\theta}{\theta}\right)^2 = 1^2 = 1$

19. $\lim_{\theta\to\pi/6} \frac{\sin(\theta-\pi/6)}{\theta-\pi/6} = \lim_{x\to 0} \frac{\sin x}{x} = 1$, on setting

$$x = \theta - \pi/6$$

and noting that this $\to 0$ as $\theta \to \pi/6$.

20. $\lim_{x\to 0} x^2 1_{\mathbb{Q}}(x) = 0$ from the 'squeeze' theorem on using
$$0 \leq \left|x^2 1_{\mathbb{Q}}(x)\right| \leq x^2 \to 0 \qquad \text{as } x \to 0.$$

21. $\lim_{x\to 0} x(1-x)1_{\mathbb{Q}}(x) = 0$ from the 'squeeze' theorem on using
$$|x(1-x)1_{\mathbb{Q}}(x)| \leq |x(1-x)| \to 0 \qquad \text{as } x \to 0.$$

22. $\lim_{x\to 1} x(1-x)1_{\mathbb{Q}}(x) = 0$ from the 'squeeze' theorem on using
$$|x(1-x)1_{\mathbb{Q}}(x)| \leq |x(1-x)| \to 0 \qquad \text{as } x \to 1.$$

23. Explain why $\lim_{x\to 3} x(x-1)1_{\mathbb{Q}}(x)$ does not exist.

Sol: Near $x = 3$ the supremum of the values $x(x-1)1_{\mathbb{Q}}(x)$ is around $3(3-1) = 6$ (actually, more than this, because if $x > 3$, with x *rational*, then $x(x-1)1_{\mathbb{Q}}(x) = x(x-1) > 3(3-1)$), whereas the inf is 0 on taking x irrational.

24. Explain why $\lim_{x\to\infty} \cos x$ does not exist.

Sol: $\cos x$ oscillates between 1 and -1 as x runs from any integer multiple of 2π (such as 0, 2π, 4π, 6π,...) and the next higher such multiple. So for any positive real number t we have
$$\sup_{x\in(t,\infty)} \cos x = \infty, \qquad \text{and} \qquad \inf_{x\in(t,\infty)} \cos x = -1.$$

Since there is no unique value lying between 1 and -1, the limit $\lim_{x\to\infty} \cos x$ does not exist.

25. Explain why $\lim_{x\to\infty} x \sin x$ does not exist.

Sol: $\sin x$ oscillates between 1 and -1 as x runs from any integer multiple of 2π (such as 0, 2π, 4π, 6π,...) and the next higher such multiple. So for any positive real number t we have

$$\sup_{x\in(t,\infty)} \sin x = \infty, \qquad \text{and} \qquad \inf_{x\in(t,\infty)} \sin x = -1.$$

Since there is no unique value lying between 1 and -1, the limit $\lim_{x\to\infty} \sin x$ does not exist.

26. Explain why $\lim_{x\to\infty} \frac{\sin x}{x} = 0$.

Sol: This follows, for instance, by the 'squeeze' theorem on observing that

$$\left|\frac{\sin x}{x}\right| \leq \left|\frac{1}{x}\right| \to 0 \qquad \text{as } x \to \infty.$$

27. Explain why $\lim_{x\to\infty} \frac{\sin x}{\sqrt{x}} = 0$.

Sol: This follows, for instance, by the 'squeeze' theorem on observing that

$$\left|\frac{\sin x}{\sqrt{x}}\right| \leq \left|\frac{1}{\sqrt{x}}\right| \to 0 \qquad \text{as } x \to \infty.$$

Solutions for Exercise Set 21.1.

(i) For the set

$$S = [-\infty, -1) \cup (1, 2] \cup \{6, 8\} \cup [9, \infty]$$

write down

(i) an interior point: -2

(ii) a limit point: -5.

(iii) a boundary point: -1

(iv) an isolated point: 6.

(v) the interior $S^0 = [-\infty, -1) \cup (1, 2) \cup (9, \infty]$

(vi) the boundary $\partial S =$

Sol: $\partial S = \{-1, 1, 2, 6, 8, 9\}$. Note that $-\infty$ and ∞ are *interior* point of S.

2. Answer and explain briefly:

(i) If $4 < \sup T$ is 4 an upper bound of T?

(ii) In (i), is there a point of T that is > 4?

(iii) If $\inf T < 3$ is 3 a lower bound of T?

Sol: Since $\inf T$ is the *greatest* lower bound of T, it follows that 3, being $> \inf T$, is *not* a lower bound of T.

(iv) In (iii), is there a point of T that is < 3?

Sol: Since 3 is *not* a lower bound of T there must be a point of T that is < 3.

3. Answer the following concerning limits, with brief explanations:

(i) If $\lim_{x\to 1} F(x) = 2$ does it follow that $F(1) = 2$?

Sol: No, the definition of the limit $\lim_{x\to 1} F(x)$ contains no information about the value of F at 1. For example,

$$F(x) = \begin{cases} 2x & \text{for all } x \neq 1; \\ 0 & \text{if } x = 1. \end{cases}$$

has $\lim_{x\to 1} F(x) = 2$ but $F(1) = 0$.

(ii) If g is continuous at 3 is g differentiable at 3?

Sol: No, a function can be continuous at a point without being differentiable at that point. For example, the function g given by

$$g(x) = |x - 3| \qquad \text{for all } x \in \mathbb{R},$$

is continuous everywhere but is not differentiable at 3.

(iii) If g is differentiable at 5 is g continuous at 5?

Sol: Yes, we have proved that if a function is differentiable at a point then it is continuous at that point.

(iv) If $h'(5) = 4$ and $h(5) = 8$ then $\lim_{x\to 5} h(x) =$

Sol: Since $h'(5) = 4$ the function h is differentiable at 5. Therefore it is continuous at 5. Hence $\lim_{x\to 5} h(x)$ is equal to $h(5)$, which is given to be 8. Thus, $\lim_{x\to 5} h(x) = 8$.

(v) If $H'(2) = 5$ and $H(2) = 3$ then $\lim_{w\to 2} \frac{H(w)-3}{w-2} =$

Sol: The limit here is

$$\lim_{w\to 2} \frac{H(w) - 3}{w - 2} = \lim_{w\to 2} \frac{H(w) - H(2)}{w - 2}.$$

We recognize this to be the derivative $H'(2)$, which is given to be 5. Thus the value of the limit is 5.

(vi) If $G'(5) = 1$ and $G(5) = 6$ then $\lim_{y\to 5} \frac{G(y)-6}{y-5} =$

Sol: The limit here is

$$\lim_{y\to 5} \frac{G(y) - 6}{y - 5} = \lim_{y\to 2} \frac{G(y) - G(5)}{y - 5}.$$

We recognize this to be the derivative $G'(5)$, which is given to be 1. TThus the value of the limit is 1.

(vii) $\lim_{w\to 0}\frac{\sin w}{w} =$

Sol: This is 1.

(viii) $\lim_{w\to \pi/3}\frac{\sin w-\sin(\pi/3)}{w-\pi/3} =$

Sol: We recognize this limit as the derivative of sin at $\pi/3$, and so its value is $\sin'(\pi/3) = \cos\pi/3 = 1/2$.

(ix) If $G'(3) = 4$ then

$$\lim_{h\to 0}\frac{G(3+h)-G(3)}{h} =$$

Sol: We recognize this limit as the derivative of G at 3 and so its value is $G'(3) = 4$, as given.

4. Work out the following derivatives:

(i) $\frac{d\sqrt{w^4-2w^2+4}}{dw}$

Sol: Using the chain rule we have

$$\frac{d\sqrt{w^4-2w^2+4}}{dw} = \frac{1}{2\sqrt{w^4-2w^2+4}}(4w^3-4w+0)$$

(ii) $\frac{d[(1+\sqrt{y})\tan y]}{dy}$

Sol: Using the product rule we have

$$\begin{aligned}\frac{d[(1+\sqrt{y})\tan y]}{dy} &= \frac{d[(1+\sqrt{y})}{dy}\tan y + (1+\sqrt{y})\frac{d\tan y}{dy}\\ &= \left(0+\frac{1}{2\sqrt{y}}\right)\tan y + (1+\sqrt{y})\sec^2 y\end{aligned}$$

(iii) $\frac{d[\frac{1+\sqrt{y}}{\tan y}]}{dy}$

Sol: Using the quotient rule we have

$$
\begin{aligned}
\frac{d[\frac{1+\sqrt{y}}{\tan y}]}{dy} &= \frac{\tan y \frac{d(1+\sqrt{y})}{dy} - (1+\sqrt{y})\frac{d\tan y}{dy}}{\tan^2 y} \\
&= \frac{\tan y(0+\frac{1}{2\sqrt{y}}) - (1+\sqrt{y})\sec^2 y}{\tan^2 y} \\
&= \frac{(\tan y)\frac{1}{2\sqrt{y}} - (1+\sqrt{y})\sec^2 y}{\tan^2 y}
\end{aligned}
$$

(iv) $\frac{d\cot x}{dx}$

Sol: Using the quotient rule we have

$$
\begin{aligned}
\frac{d\cot x}{dx} &= \frac{d\frac{\cos x}{\sin x}}{dx} \\
&= \frac{\sin x \cdot (-\sin x) \ - \ \cos x \cdot \cos x}{\sin^2 x} \\
&= \frac{-(\sin^2 x + \cos^2 x)}{\sin^2 x} \\
&= -\frac{1}{\sin^2 x} \\
&= -\csc^2 x.
\end{aligned}
$$

(v) $\frac{d\sin\left(\cos\left(\tan\left(\sqrt{x}\right)\right)\right)}{dx}$

Sol: Using the chain rule repeatedly we have

$$
\begin{aligned}
\frac{d\sin\left(\cos\left(\tan\left(\sqrt{x}\right)\right)\right)}{dx} = \cos\left(\cos\left(\tan\left(\sqrt{x}\right)\right)\right) \cdot \left[-\sin\left(\tan\left(\sqrt{x}\right)\right)\right] \\
\cdot \left[\sec^2\left(\sqrt{x}\right)\right] \cdot \frac{1}{2\sqrt{x}}
\end{aligned}
$$

5. Using the definition of the derivative, show that

$$\frac{d(1\sqrt{x})}{dx} = -\frac{1}{2x\sqrt{x}}.$$

Sol:

$$\begin{aligned}
\frac{d(1\sqrt{x})}{dx} &= \lim_{w\to x} \frac{\frac{1}{\sqrt{w}} - \frac{1}{\sqrt{x}}}{w-x} \\
&= \lim_{w\to x} \frac{\frac{\sqrt{x}-\sqrt{w}}{\sqrt{w}\sqrt{x}}}{w-x} \\
&= \lim_{w\to x} \frac{\sqrt{x}-\sqrt{w}}{(w-x)\sqrt{w}\sqrt{x}} \\
&= \lim_{w\to x} \frac{(\sqrt{x}-\sqrt{w})(\sqrt{x}+\sqrt{w})}{(w-x)\sqrt{w}\sqrt{x}(\sqrt{x}+\sqrt{w})} \\
&= lim_{w\to x} \frac{(\sqrt{x})^2-(\sqrt{w})^2}{(w-x)\sqrt{w}\sqrt{x}(\sqrt{x}+\sqrt{w})} \qquad (34.1) \\
&= \lim_{w\to x} \frac{x-w}{(w-x)\sqrt{w}\sqrt{x}(\sqrt{x}+\sqrt{w})} \\
&= \lim_{w\to x} \frac{-1}{\sqrt{w}\sqrt{x}(\sqrt{x}+\sqrt{w})} \\
&= -\frac{1}{\sqrt{x}\sqrt{x}\cdot 2\sqrt{x}} \\
&= -\frac{1}{2x\sqrt{x}}
\end{aligned}$$

Solutions for Exercise Set 20.4.

1. Find the maximum and minimum values of x^2 for $x \in [-1, 2]$.

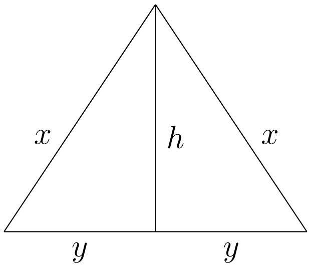
Sol: The deriative of $f(x) = x^2$ is $2x$, which is 0 at $x = 0$. Of the values $f(0) = 0$, $f(-1) = 1$, and $f(2) = 4$, the minimum value is 0 and the maximum is 4.

2. Find the maximum and minimum values of $x(6-x)(3-x)$ for $x \in [0, 2]$.

Sol: For $f(x) = x(6-x)(3-x) = x^3 - 9x^2 + 18x$, we have $f'(x) = 3x^2 - 18x + 18$, and this is 0 when $x^2 - 6x + 6 = 0$, the solutions of which are $3 \pm \sqrt{3}$ (instead of using the formula for solutions of quadratic equations it is easier to observe that $x^2 - 6x + 6 = x^2 - 2 * 3x + 9 = 3$ and this is $(x-3)^2 = 3$. Now $3 + \sqrt{3}$ is not in $[0,2]$ and $3 - \sqrt{3}$ is in $[0,2]$. We compare the values $f(0) = 0$, $f(2) = 8$, and $f(3-\sqrt{3}) = (3-\sqrt{3})(3+\sqrt{3})\sqrt{3} = 6\sqrt{3} \simeq 10.39$. So the maximum of f on $[0,2]$ is $6\sqrt{3}$ at $x = 3 - \sqrt{3}$, and the minimum is 0 at $x = 0$.

3. A wire of length 12 units is bent to form an isosceles triangle. What should the lengths of the sides of the triangle be to make its area maximum?

Sol: Let the sides be x, x, and $2y$ (it will be clear soon that it is better to take the third side to be $2y$ than y, the algebra being a bit easier). Then $2x + 2y = 12$, so $x + y = 6$. The minimum possible value of y is 0 (the triangle collapses into a line then) and the maximum value of y is 3 (again the triangle collapses into a flat line).



$$x^2 = y^2 + h^2$$

$$h = \sqrt{x^2 - y^2}$$

$$A = \frac{1}{2}2yh = y\sqrt{x^2 - y^2}$$
$$= y\sqrt{6-y)^2 - y^2} = y\sqrt{36 - 12y}$$

Thus $y \in [0,3]$. Since we have to work out the derivative of A' it will be a little easier, for algebra, to actually work the maximizing $A^2 = y^2(36 - 12y) = 36y^2 - 12y^3$ rather than A (of course, then we can just take the square root at the end). The derivative is $(A^2)' = 72y - 36y^2 = 36y(2-y)$. This is 0 at $y = 0$ and at $y = 2$, both of these being in $[0,3]$. When $y = 0$ or $y = 3$ then A is 0. So the maximum must occur when $y = 2$ and this value is $2\sqrt{36 - 12 * 2} = 4\sqrt{3}$.

4. A piece of wire is bent into a rectangle of maximum area. Show that this maximal area rectangle is a square.

Sol: Let L be the length of wire, and the suppose the rectangle has sides x and y. Then the perimeter is $2x + 2y = L$, thus $x + y = L/2$. The area is $A = xy$. Thus the problem is to maximize the product of two numbers given that their sum is $L/2$. We write A as $A = x(L/2 - x) = (L/2)x - x^2$. The restriction of x is that it lies in $[0, L/2]$. We know that the maximum value of a quadratic can be obtained without calculus, but it is simple to work out the derivative $A' = L/2 - 2x$ and observe that this is 0 when $x = L/4$. Then $y = (L/2) - (L/4) = L/4$. Thus the rectangle has equal sides, which means that it is a square. This confirms intuition.

5. A piece of wire of length L is cut into pieces of length x and $L - x$ (including the possibilty that x is 0 or L), and each piece is bent into a circle. What is the value of x which would make the total area enclosed by the pieces maximum, and what is the value of x which would make this area minimum.

Sol: If a length x is bent into a circle the radius of the circle is $x/(2\pi)$. The area is $\pi\bigl[x/(2\pi)\bigr]^2$. For the remaining length $L - x$ the area is $\pi\bigl[(L-x)/(2\pi)\bigr]^2$. Thus the total area is

$$A(x) = \pi\bigl[x/(2\pi)\bigr]^2 + \pi\bigl[(L-x)/(2\pi)\bigr]^2 = \frac{1}{4\pi}x^2 + \frac{1}{4\pi}(L-x)^2.$$

The value x lies in $[0, L]$. The derivative of A is

$$A'(x) = \frac{1}{2\pi}x + \frac{1}{2\pi}(L-x)(-1) = \frac{1}{2\pi}(2x - L).$$

This is 0 when $x = L/2$, that is, with the wire split into two equal pieces. The area enclosed is then

$$A(L/2) = \frac{1}{4\pi}(L/2)^2 + \frac{1}{4\pi}(L - L/2)^2 = \frac{L^2}{8\pi}.$$

The end point values for x are 0 and L, corresponding to rolling the wire up into just one large circle of length $L/(2\pi)$; its area is $A(0) = \frac{1}{4\pi}L^2$. Thus the minimum are is obtained by splitting the wire into two equal pieces, each rolled into a circle, and the maximum area is obtained by rolling up the entire length of wire into a circle.

6. Here are some practice problems on straight lines and distances:

(i) Work out the distance from $(1, 2)$ to the line $3x = 4y + 5$

(ii) Work out the distance from $(2, -2)$ to the line $4x - 3y - 5 = 0$.

(iii) Find the point P_0 on the line L, with equation $3x + 4y - 7 = 0$, closest to the point $(0, 3)$. What is the angle between P_0P and the line L?

(iv) Let P_0 be the point on the line L, with equation $3x+4y-11 = 0$, closest to the point $P(1, 3)$. What is the slope of the line P_oP?

(v) Let P_0 be the point on the line L, with equation $3x+4y-11 = 0$, closest to the point $P(1, 3)$. Find the equation of the line through P and P_0.

Solutions for Exercise Set 32.5.

(a) Work out the following integrals using substitutions:

i. Use

$$u = 4 - 3x \qquad \text{and} \qquad du = -3dx.$$

So then

$$dx = -\frac{1}{3}du.$$

$$\begin{aligned}\int (4 - 3x)^{2/3}\, dx &= \int u^{2/3} \left(-\frac{1}{3}du\right)\\ &= -\frac{1}{3}\int u^{2/3}\, du\\ &= -\frac{1}{3}\frac{1}{\frac{2}{3}+1}u^{\frac{2}{3}+1} + C\\ &= -\frac{1}{5}(4 - 3x)^{5/3} + C.\end{aligned} \tag{34.2}$$

ii. Use

$$y = 2 + 5x \qquad \text{and} \qquad dy = 5dx.$$

Then

$$\begin{aligned}\int \sqrt{2+5x}\,dx &= \int \sqrt{y}\,\frac{1}{5}dy \\ &= \frac{1}{5}\int y^{1/2}\,dy \\ &= \frac{1}{5}\frac{1}{\frac{1}{2}+1}y^{\frac{1}{2}+1}+C \\ &= \frac{1}{5}\frac{2}{3}y^{3/2}+C \\ &= \frac{2}{15}(2+5x)^{3/2}+C.\end{aligned} \tag{34.3}$$

iii. Use

$$w = 2-3x \qquad \text{and} \qquad dw = -3dx.$$

So

$$dx = -\frac{1}{3}dw.$$

Then

$$\begin{aligned}\int \frac{1}{\sqrt{2-3x}}\,dx &= \int \frac{1}{\sqrt{w}}\left(-\frac{1}{3}\right)\,dw \\ &= -\frac{1}{3}\int \frac{1}{\sqrt{w}}\,dw \\ &= -\frac{1}{3}\int w^{-1/2}\,dw \\ &= -\frac{1}{3}\frac{1}{-\frac{1}{2}+1}w^{-\frac{1}{2}+1}+C \\ &= -\frac{2}{3}w^{1/2}+C \\ &= -\frac{2}{3}(2-3x)^{1/2}+C.\end{aligned} \tag{34.4}$$

iv. Use

$$y = 3-2x.$$

Then

$$dy = -2dx \quad \text{and} \quad dx = -\frac{1}{2}dy.$$

$$\begin{aligned}\int x(3-2x)^{4/5}\,dx &= \int xy^{4/5}\left(-\frac{1}{2}dy\right)\\ &= -\frac{1}{2}\int \frac{3-y}{2}y^{4/5}\,dy \quad (\text{using } x=\tfrac{3-y}{2})\\ &= -\frac{1}{4}\int (3y^{4/5}-y^{9/5})\,dy\\ &= -\frac{1}{4}\left[3\frac{1}{\frac{4}{5}+1}y^{\frac{4}{5}+1}-\frac{1}{\frac{9}{5}+1}y^{\frac{9}{5}+1}\right]+C\\ &= -\frac{5}{12}y^{9/5}+\frac{5}{56}y^{7/5}+C\\ &= -\frac{5}{12}\left(3-2x\right)^{9/5}+\frac{5}{56}(3-2x)^{7/5}+C\end{aligned} \tag{34.5}$$

v. $\int \frac{x}{(2+5x)^{3/5}}\,dx$
Use the substitution

$$y = 2+5x.$$

Then

$$dy = 5dx,$$

and

$$x = \frac{1}{5}(y-2).$$

So

$$\begin{aligned}\int \frac{x}{(2+5x)^{3/5}}\,dx &= \int \frac{\frac{1}{5}(y-2)}{y^{3/5}}\,\frac{1}{5}dy\\ &= \frac{1}{25}\int \frac{y-2}{y^{3/5}}\,dy\\ &= \frac{1}{25}\int \left[\frac{y}{y^{3/5}} - \frac{2}{y^{3/5}}\right]\,dy\\ &= \frac{1}{25}\left[\int y^{2/5}\,dy - 2\int y^{-3/5}\,dy\right]\\ &= \frac{1}{25}\left[\frac{1}{\frac{2}{5}+1}y^{\frac{2}{5}+1} - 2\frac{1}{-\frac{3}{5}+1}y^{-\frac{3}{5}+1}\right]\\ &= \frac{1}{25}\left[\frac{5}{7}y^{7/5} - 5y^{2/5}\right] + C\\ &= \frac{1}{25}\left[\frac{5}{7}(2+5x)^{7/5} - 5(2+5x)^{2/5}\right] + C.\end{aligned} \tag{34.6}$$

vi. $\int \frac{2x+1}{\sqrt{x^2+x+5}}\,dx$ Use $y = x^2+x+5$, for which $dy = (2x+1)dx$ and so the integral becomes

$$\int \frac{dy}{\sqrt{y}} = 2\int \frac{dy}{2\sqrt{y}} = 2\sqrt{y} + C = 2\sqrt{x^2+x+5} + C.$$

vii. $\int e^{-x^2/2}x\,dx$

viii. Use $y = -x^2/2$, for which

$$dy = -xdx$$

and so

$$\int e^{-x^2/2}x\,dx = \int e^y(-dy) = -\int e^y\,dy = -e^y + C = -e^{-x^2/2} + C.$$

ix. $\int \frac{\sqrt{\log(2x+5)}}{2x+5}\,dx$

Use $y = \log(2x+5)$, for which

$$dy = \frac{1}{2x+5} 2dx$$

and then

$$\int \frac{\sqrt{\log(2x+5)}}{2x+5}\, dx = \int \sqrt{y}\frac{1}{2}dy = \frac{1}{2}\int y^{1/2}\, dy$$

and this equals $\frac{1}{2}\frac{1}{3/2}y^{3/2} + C$ and so

$$\int \frac{\sqrt{\log(2x+5)}}{2x+5}\, dx = \frac{1}{3}\left(\log(2x+5)\right)^{3/2} + C.$$

x. $\int \frac{1}{x\log(x)\log(\log x)}\, dx$
Using $y = \log x$ converts the integral to

$$\int \frac{1}{y\log y}\, dy$$

and then $w = \log y$ converts this to $\int \frac{1}{w}dw = \log w + C$, and so

$$\int \frac{1}{x\log(x)\log(\log x)}\, dx = \log\log\log x + C.$$

xi.

$$\begin{aligned}\int \sin(5x)\cos(2x)\, dx &= \int \frac{1}{2}\left[\sin(5x+2x) + \sin(5x-2x)\right]\, dx \\ &= \frac{1}{2}\left[-\frac{1}{7}\cos(7x) - \frac{1}{3}\cos 3x\right] + C\end{aligned} \tag{34.7}$$

xii.

$$\begin{aligned}\int \sin(5x)\sin(2x)\, dx &= \int \frac{1}{2}\left[\cos(5x-2x) - \cos(5x+2x)\right]\, dx \\ &= \frac{1}{2}\left[\frac{1}{3}\sin 3x - \frac{1}{7}\sin 7x\right] + C\end{aligned} \tag{34.8}$$

xiii. $\int \cos(5x)\cos(2x)\,dx = \frac{1}{2}\left[\frac{1}{7}\sin 7x + \frac{1}{3}\sin 3x\right] + C$

xiv. $\int \sin^3 x\,dx$

$$\begin{aligned} \sin^3 x &= \sin x \sin x \sin x \\ &= \frac{1}{2}\left[\cos 0 - \cos(2x)\right]\sin x \\ &= \frac{1}{2}\left[1 - \cos 2x\right]\sin x \\ &= \frac{1}{2}\left[\sin x - \sin x \cos 2x\right] \\ &= \frac{1}{2}\left[\sin x - \frac{1}{2}\left[\sin 3x + \sin(-x)\right]\right] \\ &= \frac{1}{2}\sin x - \frac{1}{4}\left[\sin 3x - \sin x\right] \\ &= \frac{1}{2}\sin x + \frac{1}{4}\sin x - \frac{1}{4}\sin 3x \\ &= \frac{3}{4}\sin x - \frac{1}{4}\sin 3x \end{aligned} \tag{34.9}$$

Integration gives

$$\int \sin^3 x\,dx = -\frac{3}{4}\cos x + \frac{1}{12}\cos 3x + C.$$

xv. $\int \cos^3 x\,dx$

$$\begin{aligned} \cos^3 x &= \cos x \cos x \cos x \\ &= \frac{1}{2}\left[\cos 0 + \cos(2x)\right]\cos x \\ &= \frac{1}{2}\left[1 + \cos 2x\right]\cos x \\ &= \frac{1}{2}\left[\cos x + \cos x \cos 2x\right] \\ &= \frac{1}{2}\left[\cos x + \frac{1}{2}\left[\cos 3x + \cos(-x)\right]\right] \\ &= \frac{1}{2}\cos x + \frac{1}{4}\left[\cos 3x + \cos x\right] \\ &= \frac{3}{4}\cos x + \frac{1}{4}\cos 3x \end{aligned} \tag{34.10}$$

Integration gives

$$\int \cos^3 x\,dx = \frac{3}{4}\sin x + \frac{1}{12}\sin 3x + C.$$

xvi.

$$\int \sin^2(5x)\,dx = \int \sin(5x)\sin(5x)\,dx = \int \frac{1}{2}\left[\cos 0 - \cos(5x+5x)\right]\,dx$$

and so

$$\int \sin^2(5x)\,dx = \frac{1}{2}\int \left[1-\cos(10x)\right]\,dx = \frac{1}{2}\left[x - \frac{1}{10}\sin 10x\right] + C.$$

xvii. $\int \sqrt{3-6x-x^2}\,dx$
Completing the square for x^2+6x-3 we have

$$x^2+6x-3 = x^2+2*3*x+3^2\ -3^2-3 = (x+3)^2-12$$

and so the integral is

$$\int \sqrt{12-(x+3)^2}\,dx$$

Substitute

$$x+3 = \sqrt{12}\sin\theta$$

for which

$$dx = \sqrt{12}\cos\theta\,d\theta$$

and so

$$\begin{aligned}
\int \sqrt{12-(x+3)^2}\,dx &= \int \sqrt{12-12\sin^2\theta}\,\sqrt{12}d\theta \\
&= \sqrt{12}\int \sqrt{12(1-\sin^2\theta)}\,\cos\theta\,d\theta \\
&= 12\int \cos\theta\cos\theta\,d\theta \\
&= 12\int \cos^2\theta\,d\theta = 12\int \frac{1}{2}\left[\cos 0-\cos(2\theta)\right]\,d\theta \\
&= 6\int \left[1-\cos 2\theta\right]\,d\theta \\
&= 6\left[\theta-\frac{1}{2}\sin 2\theta\right]+C \\
&= 6\left[\theta-\sin\theta\cos\theta\right]+C
\end{aligned} \tag{34.11}$$

Now substitute back in

$$\sin\theta=\frac{x+3}{\sqrt{12}} \quad \text{and} \quad \cos\theta=\sqrt{12-(x+3)^2}$$

to get

$$\int \sqrt{12-(x+3)^2}\,dx = 6\left[\sin^{-1}\frac{x+3}{\sqrt{12}}-\frac{x+3}{\sqrt{12}}\sqrt{12-(x+3)^2}\right]+C.$$

xviii. $\int \frac{1}{\sqrt{3-6x-x^2}}\,dx$

Bibliography

[1] Havil, Julian, *Gamma: Exploring Euler's Constant*. Princeton University Press (2009).

[2] Maor, Eli, *e: The Story of a Number*, Princeton University Press (2009).

[3] Nahin, Paul J., *When Least is Best: How Mathematicians Discovered Many Clever Ways to Make Things as Small (or as Large) as Possible*, Princeton University Press (2004).

[4] Napier, John, *Mirifici Logarithm* (1614).

Index

arc length, 296

absolute value
    definition, 41
    larger of x and $-x$, 42
AM-GM inequality, 212
angle
    as a pair of rays, 69
Archimedes, 12

boundary
    notation ∂S, 37
boundary points, 36

Cartesian product, 19
chain rule
    initiating examples, 137
    proof, 146
    statement, 139
chainrule
    dy/dx form, 145
closed sets, 39
    complements of open sets, 40
codomain
    of a function, 20
complement, 17
completeness
    existence of suprema, 31
    of $\mathbb{R}$, 31
    of $\mathbb{R}^*$, 31
completing squares, 153
composite function, 138
composite functions, 66
composites
    of continuous functions, 88
concave function, 206
    strictly concave, 206
continuity
    at a point, 85
continuous
    at a point, 85
    at exactly one point, 87
    at exactly two points, 88
    nowhere, 87
    on a set, ambiguity, 87
continuous functions, 86
    composites of, 88
    polynomials, 86
convergence
    of $\sum_n 1/n^2$, 259
convex combination, 210, 218
convex function, 205
    strict convexity, 206
cosecant function csc, 74
cosine
    geometric meaning, 70
cotangent function cot, 74
curve
    definition, 296

decreasing functions, 102
dense

  irrationals in $\mathbb{R}$, 51
  rationals in $\mathbb{R}$, 51
dense subset
  $\mathbb{Q}$ in $\mathbb{R}$, 26
  $\mathbb{Q}^c$ in $\mathbb{R}$, 26
derivative
  as slope of tangent line, 116
  at a point, 116
  definition, 116
  finiteness implies continuity, 131
  notation $\frac{df(x)}{dx}$, 118
  notation $df(x)/dx$, 117
  notation $f'(p)$, 116, 117
  of constant is 0, 116
  sign of, 175–179
  zero and constancy, 180
derivatives
  algebraic rules, 133
differential df, 246
differential form, 247
differential forms
  working rules, 248
discontinuity
  removable, 86
discriminant, 155
distance
  on $\mathbb{R}$, 43
  triangle inequality, 43
divergence
  of $\sum_n 1/n$, 261
domain
  of a function, 20
dummy variable, 252

elements, 13
empty set
  as subset, 16
empty set $\emptyset$, 14
extended real line $\mathbb{R}^*$, 26
exterior points, 36

factorials, 23
functions, 19
  definition, 20

graph
  of a function, 20
  of a function f, 114
  of unit circle, 22
greatest lower bound, 30

harmonic series
  divergence, 261
Hausdorff property, 35
Havil, Julian, 325

increasing functions, 102
indefinite integral, 252
indicator function, 51
infimum
  larger for larger set, 32
  notation $\inf S$, 30
  of $\emptyset$, 31
integers, 18
integrability
  and continuity, 240
integral
  of a differential, 248
  of a differential form, 253
integrand, 269
integration
  by parts, 286
  by substitution, 269
interior
  notation S^0, 37
interior point, 35
intermediate value theorem, 93

BeatriceGloria_personal library

  constructing rational powers, 95
  with intervals, 94
intersections, 17
interval, 33
inverse sin: $\sin^{-1}$ or arcsin, 100
inverse function, 99, 103
  derivative of, 204
inverse trigonometric functions, 99
irrational numbers, 18
irrationals, 26

L'Hospital's rule, 225
least upper bound, 30
length
  of a path, 295
lHospital's rule
  proof, 228
limit
  as unique value between suprema and infima, 47
  notation, 49
  of $1/x$ as $x \to 0+$, 51
  of $1/x$ as $x \to 0-$, 51
  of $1/x$ as $x \to -\infty$, 50
  of $1/x$ as $x \to \infty$, 49
limits
  'squeeze' theorem, 65
  'squeezing', 64
  and ratios, 63
  between suprema and infima, 54
  by comparision, 64
  definition with neighborhoods, 61
  of composites, 66
  of sums, 61
  products, 62
  the notion, 46
  with neighborhoods, 59
linear combination, 220
local maximum, 168
local minmum, 168
log
  and exp, 193
  as inverse exp, 193
  graph, 194
  notation ln, 193
logarithm
  definition using $1/x$, 255
lower bounds, 30
  of $\emptyset$, 30

magnitude, 41
Maor, Eli, 325
mappings
  definition, 20
maps
  definition, 20
maxima and minima
  existence, 107, 111
  with infinities, 110
Mean Value Theorem, 172
  for l'Hospital's rule, 228
minimizing quadratics, 152
monotone function, 103

Nahin, Paul J., 325
Napier, John, 325
neigborhoods, 34
neighborhoods
  and distance, 43
  of $\pm\infty$, 34

open sets, 38
  complements of closed sets, 40
  finite intersections, 39
  unions are open, 39
ordered pairs, 18

polygonal approximation
  to paths, 294
powers
  rational, definition, 196
  real, definition, 196
product rule
  and integration, 286

quadratic equations
  solutions, 154
quasi-tangents
  definition, 115
  flat at interior maxima/minima, 168
  Rolle's theorem, 172
  uniqueness and tangents, 115

radian measure, 70
Ramanujan formula, 12
range
  of a function, 21
rational numbers, 18
  notation $\mathbb{Q}$, 18
real line, 26
real numbers, 18, 26
  notation $\mathbb{R}$, 26
Riemann integral
  definition, 236
Riemann sum, 237
Rolle's Theorem, 171
Rolle's theorem
  with derivatives, 172

secan function sec , 74
secant
  to a graph, 114
semi-chord
  and sin, 71
set theory, 13
sets, 13
  equality, 14
  intersections of, 17
  unions of, 17
sin
  geometric meaning, 70
squeeze theorem, 64
strictly decreasing functions, 102
strictly increasing functions, 102
subsets, 15
  properties, 16
sum of cubes $1^3 + \cdots + N^3$, 268
sum of squares $1^2 + \cdots + N^2$, 267
summation notation, 53
supporting line, 216
  and tangent line, 216
supremum
  notation $\sup S$, 30
  of $\emptyset$, 31
  smaller for larger set, 32

tan
  geometric meaning, 70
tangent
  and maxima/minima, 168
tangent line
  and supporting line, 216
  definition, 114
triangle inequality
  for distance, 43
  for magnitudes, 42
trigonometric functions
  $\sin(1/x)$ and $\cos(1/x)$, 79
  $\sin^2 + \cos^2 = 1$, 74
  addition formulas, 75
  bounds, 77
  bounds for $(\sin x)/x$, 77
  continuity, 78

  geometric meanings, 70
  periodicity 2π, 73
  the limit $\lim_{x\to 0}\frac{\sin x}{x}=1$, 78
  values at 0, 73
  values at $\pi/2$, 73
  values at a and $-a$, 74
  values at a and $2a$, 76

uniform continuity, 240
unions, 17
upper bounds, 29
  for $\emptyset$, 29

velocity
  of a path, 293

weights, 219